UNIVERSITY OF CAPE TOWN

IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

CHARACTERISATION AND EVOLUTIONARY DYNAMICS OF TEN NOVEL *GAMMAPAPILLOMAVIRUS* TYPES FROM SOUTH AFRICAN PENILE SWABS

By

Alltalents Tutsirayi Murahwa

MPhil (MED) Immunology (UZ), Hons BMLS (UZ)

Thesis submitted to the University of Cape Town in fulfilment of the degree

Doctor of Philosophy (PhD) Medical Virology

University of Cape Town

Division of Medical Virology

Department of Pathology

Faculty of Health Sciences

University of Cape Town

January 2019

Principal Supervisor: Professor Anna-Lise Williamson

Co-Supervisor: Dr Tracy L Meiring

Preface and Declaration

All experimental work described in this thesis was carried out in the Division of Medical Virology, University of Cape Town, from February 2015 to July 2018, under the supervision of Prof Anna-Lise Williamson and Dr Tracy L Meiring.

I**, ALLTALENTS TUTSIRAYI MURAHWA,** declare that these studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma at any University. Where use has been made of the work of others, it is duly acknowledged in the text and reference. Permission of use of Figures and images from other peer-reviewed journals has been granted by the University of Cape Town, provided they are duly acknowledged in the text and reference.

Signed by candidate

Student: ALLTALENTS TUTSIRAYI MURAHWA

Dedication

To my loving parents,

Mr Cornelius Boniface Murahwa (*who passed away in March 2015 just after I started my PhD*)
and Mrs Priscah Murahwa and my son St Dominic Tawananyasha Murahwa

But it shall be one day which shall be known to the LORD, not day, nor night: ***but it shall come to pass, that at evening time it shall be light.***

Zechariah 14:7 King James Version (KJV)

Acknowledgements

To God be the glory for making this journey possible.

This project would not have been possible if it wasn't for the guidance, motivation, involvement and encouragement from my supervisors, Prof Anna-Lise Williamson and Dr Tracy Meiring, for full moral support and guidance throughout the entire project.

I would like to acknowledge colleagues from my lab for the help they rendered during my training: Zizipho Mbulawa, Ongeziwe Taku, Michel Van Diepien, Harris Onywera, Henry Munyanduki, Mani Margolin, Phindile Ximba, Warren De Moor and many others, you made my work feasible. Not forgetting a special friend, a colleague and a dearest companion Kenalemang Ntlhokoa for moral support.

Last but not least, I would like to acknowledge the Letten Foundation of Norway Oslo for financial support administered by Prof Babill Stray-Pedersen, National Research Foundation and Poliomyelitis Research Foundation for providing funding for the project and student support.

Table of Contents

Preface and Declaration ii
Dedication iii
Acknowledgements iv
Table of Contents v
List of Tables vii
List of Figures viii
List of Abbreviations and Acronyms xi
ABSTRACT 1
Chapter 1: Introduction and Literature Review 4
1.0 The *Papillomaviridae* Family 4
1.1 Nomenclature and classification of novel HPVs 5
1.2 *PAPILLOMAVIRUS* STRUCTURE AND GENOME ORGANISATION 7
1.3 *Papillomavirus* life cycle and replication 11
1.4 *Papillomavirus* conserved domains and motifs 24
1.5 Evolutionary dynamics of *papillomaviruses* 26
1.6 HPV and cancer epidemiology 31
1.7 Gammapapillomaviruses and cancer 35
1.8 HPV infections in men 38
1.9 Multiple HPV Infections 40
1.10 HPV and HIV co-infections 40
1.11 HPV genomics in the next generation sequencing era 41
1.12 Background data for this study 46
1.13 Justification 47
1.14 Aims and objectives 48
Chapter 2: Discovery, Characterisation and Genomic Variation of Ten Novel *Gammapapillomavirus* Types Isolated from Penile Swabs. 49
2.0 INTRODUCTION 49
2.1 METHODS AND MATERIALS 49
2.1.1 Ethics Statement 49
2.1.2 Penile samples 50
2.1.3 Nucleic acid isolation and amplification 50
2.1.4 Complete genome amplification 51
2.1.5 Cloning 52

2.1.6 Library preparation and sequencing 53
2.1.7 Illumina data quality control 54
2.1.8 Genomic characterisation and phylogenetic analysis 54
2.1.9 Clone selection 59
2.1.10 Nucleotide Accession numbers 59
2.1.11 Phylogenetic analysis 60
2.1.12 Nucleotide and amino acid variation 60
2.1.13 Identification of conserved domains 61
2. 2 RESULTS 61
2.2.1 Genomic organisation 61
2.2.2 Conserved domains analysis 65
2.2.3 Sequence similarity and Phylogeny 68
2.2.4 Intra-sample variations 71
2.3 DISCUSSION 87
2.4 CONCLUSIONS 94
Chapter 3: Evolutionary Dynamics of Ten Novel Gamma-PVs: Insights from Phylogenetic Incongruence, Recombination and Phylodynamic Analyses 95
3.0 INTRODUCTION 95
3.1 METHODS 96
3.1.1 Source of sequence data 96
3.1.2 Phylogenetic tree construction 96
3.1.3 Phylogenetic Incongruence tests 97
3.1.4 Recombination analysis 98
3.1.5 Construction of a time-scaled HPV phylogeny 98
3.2 RESULTS 100
3.2.1 Phylogenetic Incongruence among novel Gamma-PVs gene trees 100
3.2.2 Recombination analysis 104
3.2.3 The time-scale of Gamma-PV evolution 107
3.3 DISCUSSION 109
3.4 CONCLUSION 115
Chapter 4: General Discussion and conclusions 116
Appendix 1: Appendix to Chapter 2 120
Appendix 2: Appendix to Chapter 3 124
References 130

List of Tables

TABLE 1. 1 PVS CONSERVED DOMAINS AND MOTIF SEQUENCE PATTERNS. 25

TABLE 1. 2 INTERNATIONAL AGENCY FOR RESEARCH ON CANCER RISK CLASSIFICATION OF HPVS..... 32

TABLE 1. 3 RECENT NOVEL HPVS DISCOVERED BY NEXT GENERATION SEQUENCING. 45

TABLE 2.1 DETAILS OF PARTICIPANTS FROM WHICH NOVEL SEQUENCES WERE OBTAINED. 50

TABLE 2.2 NOVEL HPV L1 BACK TO BACK TYPE SPECIFIC PRIMER SEQUENCES..... 51

TABLE 2.3 SUMMARY OF WORK DONE ON THE AMPLIFICATION, CLONING, SANGER AND ILLUMINA SEQUENCING OF NOVEL GAMMA-HPV TYPES. 58

TABLE 2.4 GENOME LENGTHS OF NOVEL HPV TYPES AND ORF POSITIONS ON THE GENOME AND SIZES OF PROTEINS. 64

TABLE 2.5 PRESENCE AND FREQUENCY OF POTENTIAL CONSERVED DOMAINS IN THE NOVEL HPV TYPES. 67

TABLE 2.6 NUCLEOTIDE (AMINO ACID) PERCENTAGE PAIRWISE IDENTITY TO CLOSEST HPV TYPE. 69

TABLE 2.7 COMPARISON OF NUCLEOTIDE AND AMINO ACID SEQUENCE VARIABILITY AND SYNONYMOUS TO NON-SYNONYMOUS CHANGES WITHIN HPV212, HPV213 AND HPV215 GENES. 84

TABLE 2.8 COMPARISON OF NUCLEOTIDE AND AMINO ACID SEQUENCE VARIABILITY AND SYNONYMOUS TO NON-SYNONYMOUS CHANGES WITHIN HPV220 AND HPV222 GENES. 85

SUPPLEMENTARY TABLE 2.9 E1^E4 AND E2^E8 SPLICED GENE PRODUCT PREDICTION CHECKLIST. 120

TABLE 3.1 SUMMARY OF ANALYSIS DONE AND SEQUENCE DATASET USED. 99

TABLE 3.2 SHIMODAIRA-HASEGAWA TEST FOR INCONGRUENCE. .....

TABLE 3.3 INTER-SPECIES RECOMBINATION EVENT 3 IN WHOLE GENOME GAMMA-PVS. 104

TABLE 3.4 MEAN DIVERGENCE TIME OF 10 NOVEL HPVS FROM OTHER GAMMA SPECIES OR CLOSEST RELATIVE.107

List of Figures

FIGURE 1.1 HPV L1 NUCLEOTIDE SEQUENCE-BASED TAXONOMY. .....

FIGURE 1.2 HPV STRUCTURE SHOWING THE L1 PENTAMERS, TAKEN FROM (BISHOP ET AL., 2007). 8

FIGURE 1.3 HPV STRUCTURE SHOWING THE L1 IN RELATION TO L2 CAPSID PROTEINS TAKEN FROM (HAGENSEE ET AL., 1994). 8

FIGURE 1.4 HPV GENOMIC ORGANIZATION OF THE HIGH-RISK ALPHA, MU, AND BETA AND GAMMA-HPV GENOMES. 10

FIGURE 1.5 FUNCTION OF THE MAJOR GENES OF HPVS. 12

FIGURE 1.6 SCHEMATIC PRESENTATION OF THE E2 DOMAIN STRUCTURE. 14

FIGURE 1.7 (A) DIAGRAMMATIC REPRESENTATION OF HPV16 E6 SHOWING THE POSITION OF AMINO ACID MOTIFS THAT ARE IMPORTANT FOR PROTEIN INTEGRITY AND FUNCTION. (B) DIAGRAMMATIC REPRESENTATION OF E7 AND THE MOST IMPORTANT AMINO ACID MOTIFS REQUIRED FOR INTEGRITY AND PROTEIN FUNCTIONS. 19

FIGURE 1.8 HPV LIFE CYCLE AND REPLICATION. 21

FIGURE 1.9 GLOBAL SCENARIO OF PV EVOLUTION. 28

FIGURE 1. 10 TIME CALIBRATED PHYLOGENETIC TREE OF GAMMA-HPVS. 29

FIGURE 1.11 AGE STANDARDIZED WORLD INCIDENCE RATE (PER 100 000) CANCERS LINKED TO HPV, PANEL (A) ANOGENITAL CANCERS AND (B) HEAD AND NECK CANCERS (DE MARTEL, 2017). 34

FIGURE 1.12. PHYLOGENETIC TREE OF MEMBERS OF THE GENUS GAMMA-HPVS. 36

FIGURE 2.1 GENOME ANNOTATION PROCEDURE. 57

FIGURE 2.2 ALIGNMENT OF PUTATIVE E10, (37AMINO ACID PROTEIN) OF HPV214 WITH THE E10 PROTEINS OF THE MOST CLOSELY RELATED HPV TYPES FROM GAMMA-6 SPECIES 62

FIGURE 2.3 EXAMPLES OF THE GENOMIC ORGANIZATION OF THE NOVEL HPV TYPES. 63

FIGURE 2.4 ALIGNMENT OF THE E6 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 66

FIGURE 2.5 MAXIMUM LIKELIHOOD TREE OF THE NOVEL HPV TYPES AND RELATED TYPES. 70

FIGURE 2.6 HPV212, HPV213 AND HPV215 NUCLEOTIDE MISMATCH POSITIONS. 73

FIGURE 2.7 HPV220 AND HPV222 NUCLEOTIDE MISMATCH POSITIONS. 74

FIGURE 2.8 AMINO ACID VARIATIONS IN THE PREDICTED PROTEINS OF THE HPV212 CLONES (WITH CLONE 1 AS REFERENCE), HPV213 CLONES (WITH IDENTICAL CLONES 7 AND 8 AS REFERENCE CLONES) AND HPV215 CLONES (WITH CLONE 5 AS THE REFERENCE). 75

FIGURE 2. 9 AMINO ACID VARIATIONS IN THE PREDICTED PROTEINS OF THE HPV220 CLONES (WITH LONE 1 AS REFERENCE), HPV222 CLONES (WITH IDENTICAL CLONES 7 AND 8 AS REFERENCE CLONES). 76

FIGURE 2.10 HPV212 (LABELLED AS CT03 ISOLATES) CLONES ALIGNED AGAINST THE CT03 CONSENSUS SEQUENCE, HPV213 (LABELLED AS CT04 ISOLATES) CLONES ALIGNED AGAINST THE CT04 CONSENSUS SEQUENCE. .....

FIGURE 2. 11 HPV215 (LABELLED AS CT07 ISOLATES) CLONES ALIGNED AGAINST THE CT07 CONSENSUS SEQUENCE. .....

SUPPLEMENTARY FIGURE 2.12 ALIGNMENT OF THE E7 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 120

SUPPLEMENTARY FIGURE 2.13 ALIGNMENT OF THE E1 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 121

SUPPLEMENTARY FIGURE 2.14 ALIGNMENT OF THE E1 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 121

SUPPLEMENTARY FIGURE 2.15 ALIGNMENT OF THE L1 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 122

SUPPLEMENTARY FIGURE 2.16 ALIGNMENT OF THE E2 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 122

SUPPLEMENTARY FIGURE 2.17 ALIGNMENT OF THE L2 PROTEINS OF THE NOVEL HPVS AND CLOSELY RELATED HPVS. 123

FIGURE 3. 1 MINIMUM BRANCH LENGTH DISTANCE (K TREE SCORE) BETWEEN PHYLOGENETIC TREES CONSTRUCTED USING DIFFERENT HPV GENES. 101

FIGURE 3. 2 A REPRESENTATIVE PHYLOGENETIC INCONGRUENCE BETWEEN EARLY AND LATE GENES AMONG THE 10 NOVEL TYPES AND THEIR CLOSEST KNOWN GAMMA-PVS. 102

FIGURE 3. 3 FAST NJ TREE FROM A) MAJOR PARENT (1-4445 AND 8019-9792) AND FROM B) MINOR PARENT (4446-8018). **106**

FIGURE 3. 4 MOLECULAR DIVERGENCE TIMES OF PVS. .....

SUPPLEMENTARY FIGURE 3. 5 POSTERIOR SUPPORT VALUES OF PV DIVERGENCE ESTIMATES. .....

SUPPLEMENTARY FIGURE 3. 6A- F (BELOW) SIX TABLES SHOWING THE SHIMODAIRA-HASEGAWA TEST RESULTS USING W-IQ-TREE. **126**

List of Abbreviations and Acronyms

%	Percent
˚C	Degrees Celsius
®	Registered
µg	Micrograms
µl	Microliters
µm	Micrometres
aa	amino acid
AIDS	Acquired Immunodeficiency Syndrome
Alpha-PV	*Alphapapillomavirus*
aLRT	Approximate likelihood ratio test
APC	Antigen-presenting cell
APOBEC	Apolipoprotein B mRNA Editing Catalytic Polypeptide
ATP	Adenosine triphosphate
BEAST	Bayesian evolutionary analysis sampling tress
Beta-PV	*Betapapillomavirus*
BLD	Branch length distances
BM	Basement membrane
bp	Base pair
Brd4	Bromodomain-containing protein 4
CcPV1	*Caretta caretta PV1*
CD4	Cluster of Differentiation antigen number 4
CD4+	CD 4 positive effector T lymphocytes
CDD	Conserved domains database
CTL	Cytotoxic T lymphocyte
DFKZ	German Cancer Research Centre
DNA	Deoxyribonucleic acid
EGFR	Epidermal growth factor receptors
EV	Epidermodysplasia verruciformis
FcPV	*Fringilla coelebs PV*
Gamma-HPV	*Gammapapillomavirus*
GTR	General Time Reversible
HIM study	HPV in Men study
HIV	Human immunodeficiency virus
HPD	Highest posterior density
HPV	Human papillomavirus
HREC	Health Research Ethics Committee
hrHPV	High-risk HPV types
HSPG	Heparin sulfate proteoglycans
IARC	International Agency for Research on Cancer
ICC	Invasive cervical cancer
ICTV	International committee on taxonomy of viruses
IFN-γ	interferon gamma
kbp	Kilobase pair
kDA	KiloDalton
KGFR	Keratinocyte growth factor receptors
LCA	Last common ancestor
LCR	Long control region
LPS	Lipopolysaccharide
LTNPs	Long-term non-progressors
MCC	Maximum Clade Credibility
MCMC	Markov Chain Monte Carlo
MHC	Major histocompatibility Complex
ml	Millilitre
mM	Millimolar
MRCA	Most recent common ancestor
mRNA	Messenger RNA
MSM	Men who have sex with men
MSW	Men who have sex with women
Mu-PV	*Mupapillomavirus*
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
MYA	Million years ago
NES	Nuclear export signal
NGS	Next generation sequencing
NIAID	National Institute of Allergy and Infectious Diseases
NLS	Nuclear localisation signal
nm	Nanometre
nt	nucleotide
Nu-PVs	*Nupapillomavirus*
OD	Optical density
ORF	Open reading frame
PaVE	Papillomavirus Episteme
PBMC	Peripheral blood mononuclear cells

PBS	Phosphate buffered saline
PBM	PDZ Binding Motif
PCR	Polymerase chain reaction
PDGF	Platelet derived growth factor
PePV	*Psittacus erithacus PV*
pfu	Plaque forming units
pg	Picogram
pRB	Retinoblastoma proteins
PV	*Papillomavirus*
RDP	Recombination detection program
RT	Reverse transcriptase
SCC	Squamous cell carcinoma
SH test	Shimodaira-Hasegawa test
SNIP	Single nucleotide polymorphism
SPRI	Solid phase reversible immobilisation
TBP	TATA-binding protein
UCT	University of Cape Town

ABSTRACT

Human papillomaviruses (HPVs) are genetically diverse, belonging to five distinct genera: Alpha, Beta, Gamma, Mu and Nu. We discovered ten novel *Gammapapillomaviruses* (Gamma-HPVs). Genomic characterisation and phylogenetic evaluation of the ten novel Gamma-HPV types were done: HPV211, HPV212, HPV213, HPV214, HPV215, HPV216, HPV219, HPV220, HPV221 and HPV222. These HPVs were previously identified in a study that was done on 218 penile samples (104 HIV negative and 114 HIV positive) using high throughput sequencing (Roche 454) of amplimers obtained using FAP59/64 primers which were designed to detect "cutaneous" or Beta- and Gamma-HPVs. Fifteen putative novel HPV types were identified from the short HPV L1 FAP fragments HPV211 (CT02, KY063000), HPV212 (CT03, KY063001), HPV213 (CT04, KY063002), HPV214 (CT06, KY063004), HPV215 (CT07, KY063005), HPV216 (CT12, KY063010), HPV219 (CT01, KY062999), HPV220 (CT08, KY063006), HPV221 (CT09, KY063007) and HPV222 (CT155, AY009886) with prevalences varying from 0.5% to 4.1% of men sampled.

Multiple full genome clones for each novel type were generated through whole genome amplification, cloning and next generation sequencing. Complete genome sizes were: HPV211 (7253 bp), HPV212 (7208 bp), HPV213 (7096 bp), HPV214 (7357 bp), HPV215 (7186 bp), HPV216 (7233 bp), HPV219 (7108 bp), HPV220 (7381 bp), HPV221 (7326 bp) and HPV222 (7275 bp). Phylogenetically the novel *Papillomaviruses* (PVs) all clustered with Gamma-HPVs: HPV211 is most closely related to HPV168 (72% identity in the L1 nucleotide sequence) of the Gamma-8 species, HPV212 is most closely related to HPV144 (82.9%) of the Gamma-17 species, HPV213 is most closely related to HPV153 (71.8%) of the Gamma-13

species, HPV214 is most closely related to HPV103 (75.3%) of the Gamma-6 species, HPV215 and HPV216 are most closely related to HPV129 (76.8% and 79.2% respectively) of the Gamma-9 species. HPV219 is phylogenetically most closely related to HPV213 (87% identity in L1 gene) of the Gamma-13 species, HPV220 to HPV212 (72%) of Gamma-17, HPV221 to HPV142 (80%) of Gamma-10, HPV222 to HPV162 (73%) of Gamma-19.

The novel HPV types demonstrated the classical genomic organisation of Gamma-HPVs, with seven open reading frames (ORFs) encoding five early (E1, E2, E4, E6 and E7) and two late (L1 and L2) proteins. Typical of Gamma-HPVs, the novel types all lacked the E5 ORF and HPV214 also lacked the E6 ORF. We further examined variation of the novel types in clinical specimens from which they were identified. All the clones of HPV211, HPV214, HPV216, HPV219 and HPV221 were identical and showed 100% pairwise identity. The clones of HPV213, HPV215, HPV212, HPV220 and HPV222 had several differences. Analysis of mismatches between the nine genomic clones of HPV212 showed a total of 67 mismatch positions that varied along the 7208 bp genome and all the clones were unique. Analysis of mismatches between the 10 genomic clones of HPV213 showed a total of 51 mismatch positions that varied along the 7096 bp genome and it had 5 unique clones. The 6 genomic clones of HPV215 showed a total of 50 mismatch positions along a 7186 bp genome and it had 3 identical and 3 different clones. HPV220 had 4 different genomic clones that showed 17 mismatch positions along a 7381 bp genome. The 5 different clones of HPV222 showed a total of 24 mismatch positions along the 7275 bp genome. Conserved domains observed among the novel types were the Zinc finger binding Domain and PDZ domains. A retinoblastoma binding protein (pRB) binding domain in the E7 protein was additionally identified in HPV214 and HPV222. PVs are thought to evolve slowly because they co-opt

high-fidelity host cellular DNA polymerases for their replication. Despite extensive efforts to catalogue all the HPV species that infect humans, it is likely that many still remain undiscovered. We used the genome sequences of the ten novel viruses and related HPVs to analyse the evolutionary dynamics of these viruses at the whole genome and individual gene scales. We found statistically significant incongruences between the phylogenetic trees of different genes which imply gene-to-gene variation in the evolutionary processes underlying the diversification of Gamma-PVs. We were, however, only able to detect convincing evidence of a single recombination event which, on its own, cannot explain the observed incongruences between gene phylogenies. The divergence times of the last common ancestor (LCA) of the Alpha, Beta, Mu, Nu and Gamma genera was predicted to have existed between 49.7-58.5 million years ago before splitting into the five main lineages. The LCA of the presently sampled Gamma-PVs was predicted to have existed between 45.3 and 67.5 million years ago: approximately at the time when the simian and tarsier lineages of the primates diverged. The discovery, characterisation and classification of HPV211, HPV212, HPV213, HPV214, HPV215 HPV216, HPV219, HPV220, HPV221 and HPV222 add these novel types to the repertoire of the ever expanding Gamma-HPVs genus hence expanding our knowledge of these viruses.

Chapter 1: Introduction and Literature Review

1.0 THE *PAPILLOMAVIRIDAE* FAMILY

Human papillomavirus (HPV) is a member of the *Papillomaviridae* family (Bernard et al., 2010, de Villiers et al., 2004) which was once part of the larger family of *Papovaviridae* which was split into *Polyomaviridae* and *Papillomaviridae* by the International Committee on Taxonomy of Viruses (vanRegenmortel, 2002). " According to the most recent ICTV classification, the *Papillomaviridae* family includes two subfamilies *Firstpapilomavirinae* with more than 50 genera and *Secondpapillomavirinae* with one genus and one species"(Van Doorslaer et al., 2018). Genera are named according to the Greek alphabet from alpha to omega and following exhaustion of the alphabet the term dyo- and treis- (Greek for second time and third respectively) coined to accommodate the extra genera e.g. dyo-*deltapapillomaviruses* (Bernard et al., 2010). HPVs are distributed over 5 genera (Alpha, Beta, Gamma, Mu and Nu). The other papillomavirus (PV) genera are from other mammals (2007.), birds (3) and reptiles (1) (Bernard et al., 2010). Below the genus level are species and below the species level are types (de Villiers et al., 2004). Different genera have less than 60% similarity within the L1 gene, while species share between 60 and 70% similarity and types share types share between 71% and 89% similarity. The ICTV is responsible for nomenclature of viruses down to species level, and below species level, the International HPV Reference Centre assigns unique HPV type numbers after the complete genome has been sequenced, cloned and confirmed by the Centre (Mühr et al., 2018).

1.1 NOMENCLATURE AND CLASSIFICATION OF NOVEL HPVS

The recognition of a novel HPV type by the International HPV Reference Centre and the scientific community is based on availability of the full cloned genome, with the L1 gene sequence greater than 10% different or <90% similar from any previously described type (Bernard et al., 2010) (Figure 1.1).

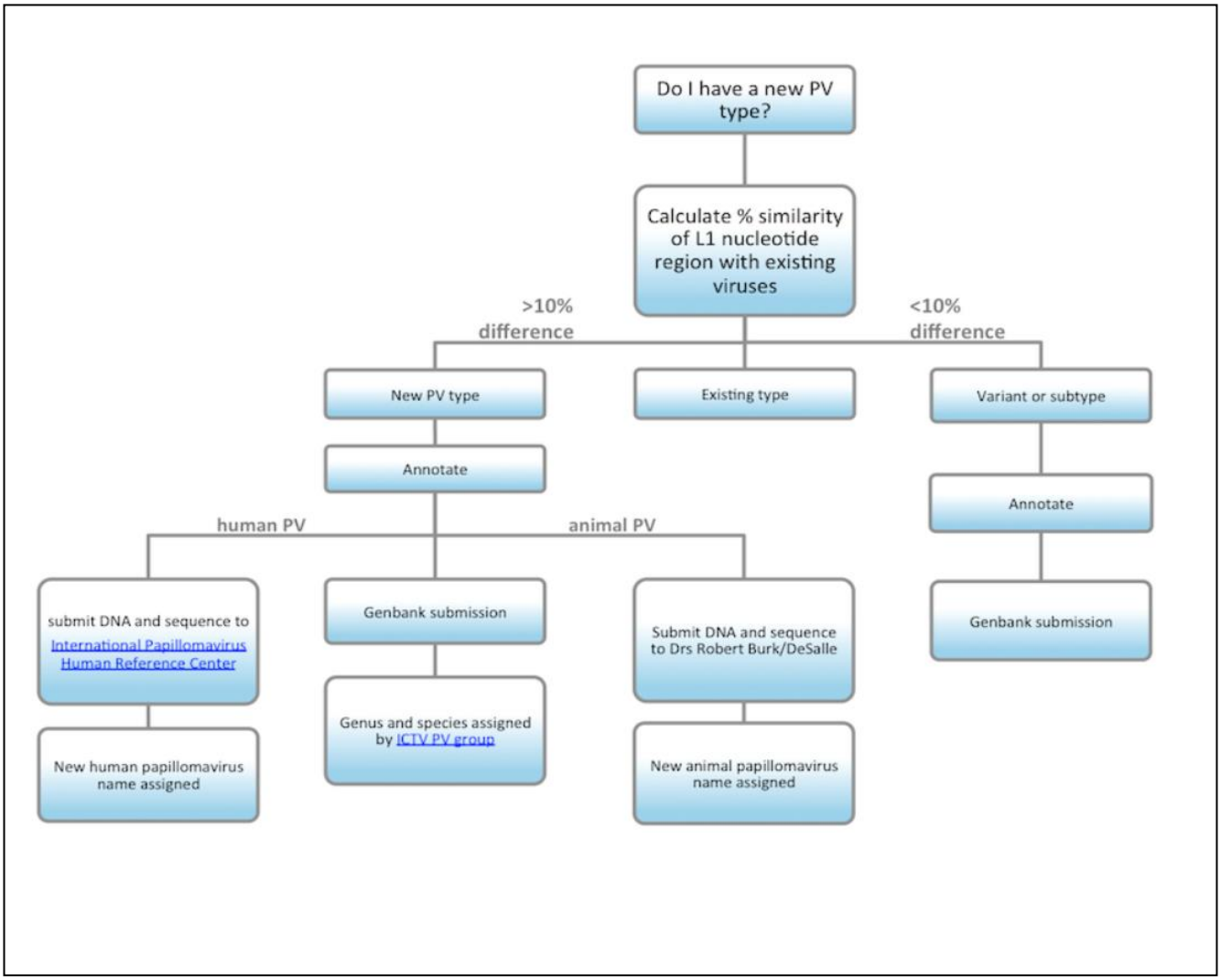


Figure 1.1 HPV L1 nucleotide sequence-based taxonomy.

https://pave.niaid.nih.gov/#explore/taxonomy/submission_process

On observation of a potential novel HPV type, the DNA has to be amplified to obtain the complete genome; cloned into a plasmid; sequenced, submitted for official naming; classified into the correct phylogenetic position and have its genes annotated (Chen et al., 2015). Determination of sequence similarity and differences requires a rigorous interrogation of the potential HPV type with sequences of all other HPV types in various databases (Harari et al., 2014).

In 1985, an International HPV Reference Centre was established at German Cancer Research Centre (DFKZ) in Heidelberg and transferred to the Karolinska Institute (Stockholm) in 2012. The International HPV Reference Centre has to date discovered about 360 new previously unknown HPV types (Arroyo Muhr et al., 2014, Ekstrom et al., 2013b) and postulates the total number of HPV types to be in the region of 400 (Bzhalava et al., 2015). Since the transfer of the HPV reference laboratory from the DFKZ to the International HPV Reference Centre in Karolinska Sweden in 2012, a total of 55 novel HPV types have attained number designation status, bringing the highest number to HPV226 (Latsuzbaia et al., 2018). And since 2014, 23 novel types have been established, 19 (82.6%) of which belong to the Gamma genus (Mühr et al., 2018). Essentially, there are 221 HPV types due to a re-classification of 4 previously known types i.e. HPV46 reclassified as HPV20, HPV55 as HPV44, HPV64 as HPV34 and HPV79 replaced by HPV91 (Terai and Burk, 2002) as these did not meet the updated criteria as unique HPV types. The 221 HPV types fall in to 5 different genera: Alpha (65 types), Beta (54 types), Gamma (98 types), Mu (3 types) and Nu (1 type).

We report in the second chapter of this thesis the discovery and characterisation of ten novel *Gammapapillomaviruses* (Gamma-HPVs) and thus adding to the list of the ever-expanding genus. The addition of more PV sequences has a bearing on the understanding of the origin, evolution and clinical outcome prediction of given PV genomes. The proposition to include the four backbone ORFs that all PVs possess (E1, E2, L1 and L2) in the re-classification of PVs, will be discussed briefly in the same chapter.

1.2 *PAPILLOMAVIRUS* STRUCTURE AND GENOME ORGANISATION

Structure: PV particles all share a similar non-enveloped icosahedral structure (50-60 nm in diameter) irrespective of genus (Doorbar et al., 2015). The structure of the virus coat consists of 360 molecules of L1 arranged into 72 capsomers, each capsomer is made up of 5 L1 molecules (Schiller and Lowy, 2012, Modis et al., 2002, Chen et al., 2000, Hagensee et al., 1994). Capsomers interact through disulphide bonds at the C-terminal ends of the L1 protein (Wolf et al., 2010, Schiller and Lowy, 2012) [Figure 1.2]. Therefore, each virion is composed of an outer shell of 72 capsomers comprising 60 hexameric and 12 pentameric capsomers of the major capsid protein (L1) arranged on a T=7 dextrose lattice (Zhou et al., 1993, Conway and Meyers, 2009). The T=7 lattice cannot be obtained for visualisation from *in vitro* cultures, instead a T=1 lattice (Figure 1.2) can be visualised. The T=7 lattice is the speculated whole structure of the virion particle from X-ray crystallographic extrapolations (Bishop et al., 2007), [Figure 1.3]. The minor capsid protein (L2) is arranged in the inner side of the virus particle under the L1 protein, it is between the L1 and L2 capsid proteins that the circular DNA is packaged as a mini-chromosome (Modis et al., 2002). Thus, the L1 and L2 proteins form the overall capsid of the viral particle and antigenic differences in these have been shown between different HPV strains (Bordeaux et al., 2006, Carter et al., 2006).

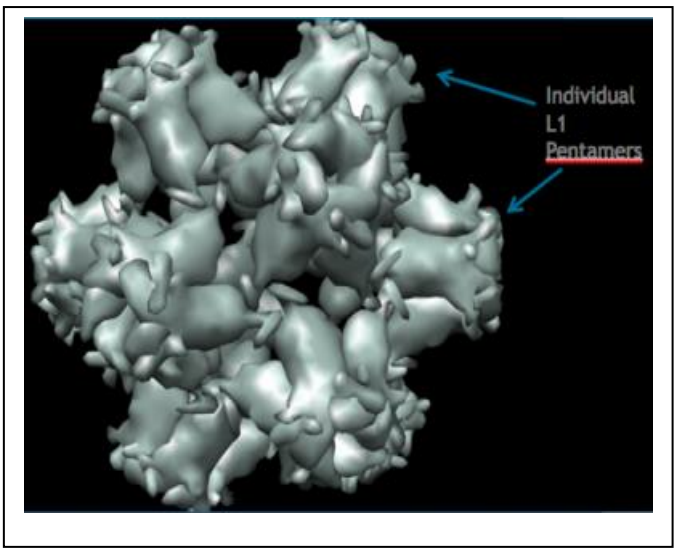


Figure 1.2 HPV structure showing the L1 pentamers, taken from (Bishop et al., 2007).

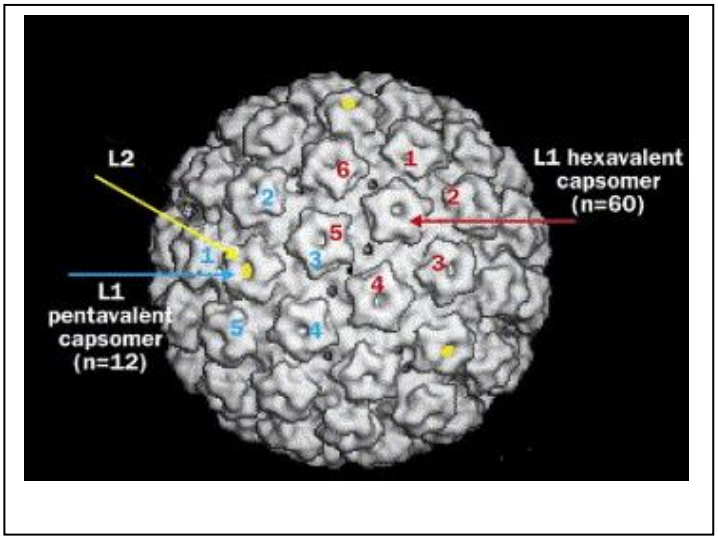


Figure 1.3 HPV structure showing the L1 in relation to L2 capsid proteins taken from (Hagensee et al., 1994).

PV virion particles contain variable numbers of L2 molecules, with only portions of the L2 being exposed to the outer surface (Rubio et al., 2011, Liu et al., 1997). The major surface-exposed protein is the L1, which comprises a series of hypervariable amino acid regions or loops that have diverged between different PV types, due to host immune selection pressure, with neutralising antibodies raised to one HPV type binding to distantly related types only poorly (Wang and Roden, 2013). This varied cross-protection has obvious practical consequences in vaccine development.

Genome: The PV genomes comprise of double-stranded circles of approximately 8 kbp in size containing 9-10 ORFs. The 8 most common genes are E1, E2, E4, E5, E6, E7, L1 and L2 (Syrjanen, 2018). The key proteins encoded are either structural (L1 and L2) or are involved in DNA replication (E1 and E2), while the other proteins (E4, E5, E6 and E7) are not necessarily encoded by all PVs and are considered as evolutionary modifications (McBride, 2017). Due to the small size of the PV genome and in order to fulfil the requirements of viral replication and survival, the genome allows for double coding and alternative protein isoforms (http://pave.niaid.nih.gov/#home 24) which is achieved through gene splicing. The circular double-stranded DNA genome is organised into three major regions: (i) an upstream regulatory region (Li et al., 2012), also known as the long control region (LCR), that is the origin of replication and houses transcription factor-binding sites and regulates gene expression; (ii) an early region, encoding for six genes involved in various functions ranging from viral replication to cell transformation (E1, E2, E4, E5, E6, E7), and (iii) a late region, encoding for the L1 and L2 structural capsid proteins which self-assemble to yield the virion particle (Bravo and Felez-Sanchez, 2015). *Betapapillomaviruses* (Beta-HPVs) and Gamma-

HPVs lack the E5 ORF, which in the *Alphapapillomaviruses* (Alpha-HPVs) genera is located downstream of the E2 ORF (Figure 1.4).

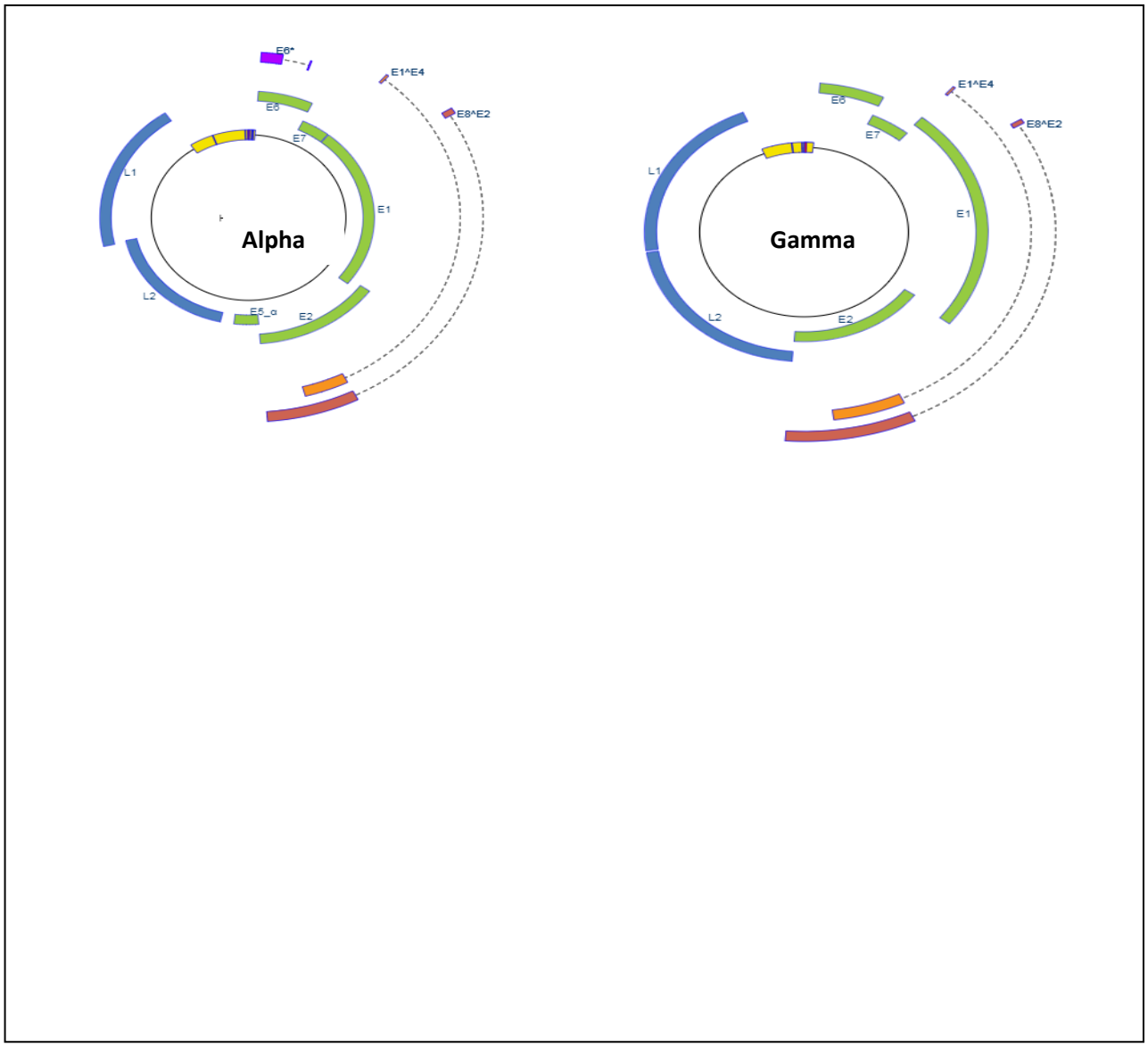


Figure 1.4 HPV genomic organization of the high-risk Alpha, Mu, and Beta and Gamma-HPV genomes.

Although all share a common genetic organization, the size and position of the major ORFs can vary, with Beta, Gamma and Mu- HPV types lacking an E5 ORF. Adapted with amendments from PaVE database https://pave.niaid.nih.gov/ (accessed 27/12/2018).

The E4 protein (whose role is in virus escape from the epithelial surface) shows the greatest sequence heterogeneity among different HPV types. This E4 heterogeneity is thought to reflect the different tissue tropisms, probably host specificity and transmission routes of different PVs (DiMaio and Petti, 2013, Doorbar, 2013). Perhaps more surprisingly, given its importance in genome amplification, is the absence of the HPV E6 protein in other Gamma-HPV species, which will be discussed in detail later.

1.3 *PAPILLOMAVIRUS* LIFE CYCLE AND REPLICATION

The most widely studied model of HPV life cycle and replication is that of HPV16 due to its involvement in anogenital and other cancers (Forman et al., 2012). Although this may be applicable to other HPVs, differences do exist between types, species and genera. Described here is the life cycle of HPV16 representing that of high-risk HPV types (hrHPV), which are associated with malignancy. The low-risk PV types, which are rarely found in cancers, differ from high-risk types in a variety of ways ranging from genome content and organisation, to gene functionality to biological and biochemical activities (Egawa and Doorbar, 2017). Of the low-risk types, the Alpha low-risk types HPV11 (Goon et al., 2008) and HPV6 (Cornall et al., 2013) have been studied, though not adequately. Of the Beta types, HPV5 and HPV8 have also been studied and associated with Epidermodysplasia verruciformis (EV) (Lazarczyk et al., 2009), while for the Gamma types there is very little information on their life cycle as they have not been shown to be associated with important cancers.

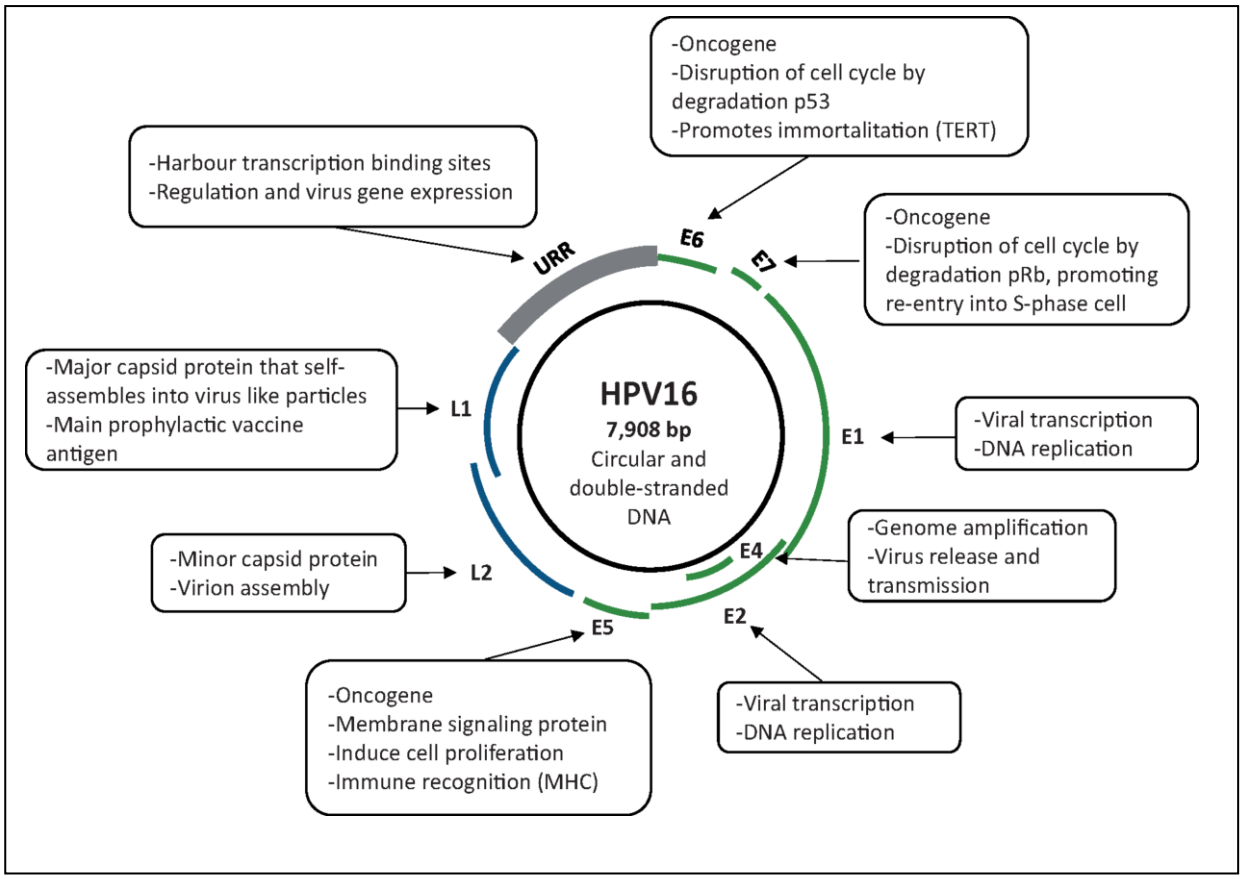


Figure 1.5 Function of the major genes of HPVs.

HPV genome is divided into three major regions: the LCR, an early region encoding six genes (E1, E2, E4, E5, E6, E7) and a late region encoding for the L1 and L2 capsid proteins. Taken from (Taberna et al., 2017) **.**

HPV cell entry: PV virions enter the basal keratinocytes either through abrasion, small wounds or hair follicles (Doorbar, 2005). It has been shown that heparin sulfate proteoglycans (HSPG) play a role in virion entry into cells through interacting with L1 capsid proteins (Schiller et al., 2010, Aksoy et al., 2017) [Figure 1.8]. Receptor binding mechanisms maybe dependant on the HPV type or the cell type being infected or that multiple receptor strategies maybe involved in a single infection (DiGiuseppe et al., 2017). Important to note is that with HPV, cell attachment and entry into basal keratinocytes are not the rate limiting steps to infection (Handisurya et al., 2012) because infection does not require an intact virion, the naked viral DNA is equally infective (Bravo and Felez-Sanchez, 2015). Several

studies have shown that HSPG are the not the sole receptors, but several other receptors have been identified through experiments done mainly using HPV16 (Selinka et al., 2007, Day et al., 2008). Among the proteins that have been shown to be involved in viral entry are: α6 and β4 integrins (Evander et al., 1997), tetraspanins CD63 and CD151 (Spoden et al., 2008), annexin A2 heterotetramer (Woodham et al., 2012) and epidermal growth factor receptors/ Keratinocyte growth factor receptors (EGFR/KGFR) (Dziduszko and Ozbun, 2013).

Upon cell binding, the virus needs to be internalised into the cells in order to deliver the DNA to the host nucleus for a productive infection (Aksoy et al., 2017). There are two well described pathways for endocytosis namely the clathrin and caveolin pathway, all common among non-enveloped viruses (Pelkmans and Helenius, 2003). The mechanism of endocytosis used by both pathways is based on dynamin (a guanine triphosphatase enzyme responsible for endocytosis in eukaryotic cells), which through its hydrolase activity pinches off endocytic vesicles from the plasma membrane (Aksoy et al., 2017). Work done using HPV16 endocytosis showed that it used the clathrin mediated endocytosis pathway (Smith et al., 2007) and it was later also shown that both the clathrin and caveolin mediated pathways play a role in cell entry (Laniosz et al., 2009). There have not been any similar studies on other HPV genera save for Alpha-HPVs.

Early phase of viral replication cycle: E1 and E2: The E1 protein contains three important domains, the ATP binding site, a bipartite nuclear localisation signal (NLS) and a nuclear export signal (NES). NES and NLS motifs are short sequence motifs (Table 1.1) which facilitate protein transport in and out of cell membranes (Bernhofer et al., 2018). NLS motifs vary in length but all share a simple consensus sequence K-K/R-X-K/R, and if the motif is a single one, it is a monopartite and if they are two it is bipartite (Bernhofer et

al., 2018), as shown for HPVs in the Table 1.1. The NES was first described in HIV-1 (Wen et al., 1995) and attempts to define a consensus sequence have been proposed but much variation was observed (la Cour et al., 2004). The classical NES motif however has 3-4 hydrophobic amino acids especially leucine (Fischer et al., 1995). The ATP binding domain of the E1 is located in the C-terminus of the protein. The ATP binding domain is important for recruitment of E1 to the origin of replication in the LCR (Titolo et al., 1999).

The E2 protein has two main domains: a NLS and DNA recognition helix. The NLS is as described in the E1 protein save that for E2, it is a monopartite domain in the transactivation domain as shown in Table 1.1. The DNA recognition helix of E2 is in the DNA binding domain as shown in Figure 1.6 below.

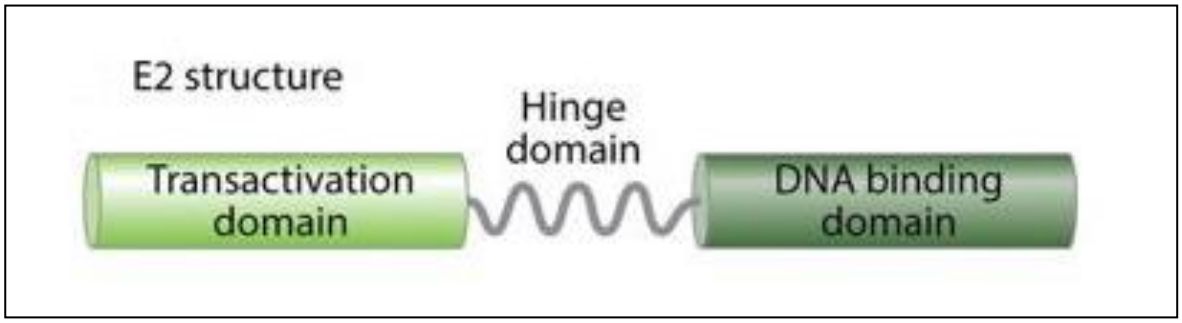


Figure 1.6 Schematic presentation of the E2 domain structure.
The transactivation domain- Light green; the DNA binding domain dark green; the hinge domain- gray wavy line. Taken from (Graham, 2016).

PVs have acquired a unique ability to take advantage of the self-renewal process of stratified cutaneous and mucosal epithelia, as they replicate in synchronisation with epithelial differentiation from basal keratinocytes to stratum spinosium and to stratum granulosum (McBride, 2017). Viral replication is performed by high fidelity host DNA

polymerases in parallel with host DNA replication (Doorbar et al., 2012). Upon nuclear entry into the dividing epithelial cells, the first goal of the virus is to carry out the initial replication of its genome, this is achieved by replication and transcription factors of E1 and E2 (Graham, 2017): the early expression of the viral transcription factor E1 allows for the regulation of early viral promoter to modulate expression of E6 and E7 regulatory proteins that ensures survival of HPV infected epithelial cells (Moody and Laimins, 2010). The interplay between E2 and E1, during early replication forms the basis of initial successful replication: the E2 protein consists of two domains: a DNA binding domain and a protein binding domain linked together by a flexible hinge domain (van de Poel et al., 2018). The E2 forms a homodimer that can bind to four palindromic sites in the LCR with three of these sites situated next to the origin of replication and these three sites in the LCR are reserved for E1 mediated replication (McBride, 2013).

Consequently, E2 binds E1 which in turn through a dimer of hexamers binds to the viral origin of replication in the LCR and thus activating cellular DNA replication machinery (McBride, 2017, Bergvall et al., 2013b). After this initial infection, incoming HPV genomes are replicated to produce around 50 to 100 episomal copies per cell, and this is the first phase of viral replication (Maglennon et al., 2011). The infected cell then leaves the basal layers and enters the transit enhancing proliferative layers of the epithelium where there is episomal and plasmid maintenance of virion particle numbers (Pinidis et al., 2016). In the infected basal cells, the HPV circular DNA is replicated parallel to the host DNA and in turn partitioned to daughter epithelial cells at differentiation. The partitioning into daughter cells is achieved through tethering (binding) of viral genomes to host cell chromosomes via E2 that is bound to the LCR of the virus and other chromatin binding proteins (McBride, 2013).

The most widely studied anchor of viral genomes to host chromosomes is Bromodomain-containing protein 4 (Brd4) that works in concert with E2 (Iftner et al., 2017), but many other alternative anchors have been studied. So HPV has mastered the art of carrying out its replication cycle in unison with epithelial cell differentiation and evading immune surveillance (Stanley, 2012), this is achieved through a series of a well-timed program of gene expression (Egawa et al., 2015).

E4, E5, E6 and E7 protein activities during replication: The biological roles of E6 and E7 are essential in the early replication stages in the basal epithelial layers. The E6 and E7 proteins disrupt the checkpoint mechanisms of the cell cycle making sure all stages are completed properly, thus facilitating uncontrolled keratinocyte differentiation. PDZ proteins is a family of proteins containing the PDZ domain. PDZ is an abbreviation combining the first three proteins discovered to have the domain: post synaptic density protein (PSD95), Drosophila disc large tumour suppressor (Dlg1), and zonula occludens-1 protein (zo-1) (Kennedy, 1995). Binding to PDZ motifs is mediated by short peptide sequences referred to here as PDZ Binding Motif (PBM), usually located at the C-terminus of proteins (James and Roberts, 2016). The PDZ domain-binding motif -X-(T/S)-X-(L/V) at the carboxy end of E6 is important for targeting PDZ proteins for proteasomal degradation, see Table 1.1. PBM binds to a whole range of other proteins giving E6 an essential role in viral life cycle and survival. The E6 PBM is required for maintenance of episomal viral DNA and genome amplification usually in high risk HPVs (Delury et al., 2013). The E7 protein binds to pRB and degrades them resulting in cell cycle dysregulation (Roman and Munger, 2013a). The E6 protein also prevents apoptosis through its binding of p53 tumour suppressor (Martinez-Zapien et al., 2016).

The E6 protein is about 150 amino acids long and about 18 kDa (Bedard et al., 2008). The E6 has two main conserved domains, two Zinc fingers and a PBM (Miller et al., 2012) [figure 1.7A]. Zinc fingers are small structural protein motifs with finger-like protrusions, and Zinc finger domains are protein structures that contain multiple finger-like protrusions (Laity et al., 2001). Zinc finger domains are important structures in the synthesis of fusion protein complexes and serve a wide variety of biological functions (Roman and Munger, 2013b). The E7 of most HPVs have one Zinc finger domain and may have a pRB binding domain (Figure 1.7B). The structure and function of the Zinc finger domain is as described for the E6. The pRB binding domain has a short peptide sequence signature (Table 1.1) as previously described, this domain is in part responsible for HPV carcinogenicity through its binding of the pRB (Roman and Munger, 2013a).

HPV E5 proteins are small, transmembrane based and hydrophobic proteins spanning about 40 to 93 amino acids in size (DiMaio and Petti, 2013). The E5 ORF is classified into four groups (alpha, beta, gamma and delta) based on clinical manifestation, particularly oncogenic potential (Bravo and Alonso, 2004a, Schiffman et al., 2005b). High risk HPVs encode the E5-alpha while E5-gamma and delta are encoded for by low risk HPVs (Garcia-Vallve et al., 2005).

As previously alluded to, the E5 is absent in many HPV genomes and hence pointing to the fact that it is not essential for the HPV life cycle but gives Alpha-PVs an added advantage to favour infection and transformation (Venuti et al., 2011). The E5 protein favours the overgrowth and prevents apoptosis of infected cells and hence induces malignant progression (DiMaio and Petti, 2013). E5 also helps the virus in evading immune response

through repression of major histocompatibility complex (MHC) presentation of viral peptides

(Ashrafi et al., 2006). The transforming activity of the E5 has been demonstrated in various cell types (Ghai et al., 1996, Rho et al., 1996). Experiments with HPV6 provided the first evidence of E5 transforming activity in mammalian cells where it was shown that the expression of HPV6 E5 led to the anchorage and independent growth of murine fibroblasts (Chen and Mounts, 1990). Subsequently, it was also shown that HPV16 E5 induces anchorage and independent growth of murine keratinocytes and fibroblasts (Leechanachai et al., 1992, Pim et al., 1992). Thus, HPV E5 proteins are not strongly carcinogenic and not directly important for carcinogenesis because they are not always found in HPV associated tumours (Venuti et al., 2011).

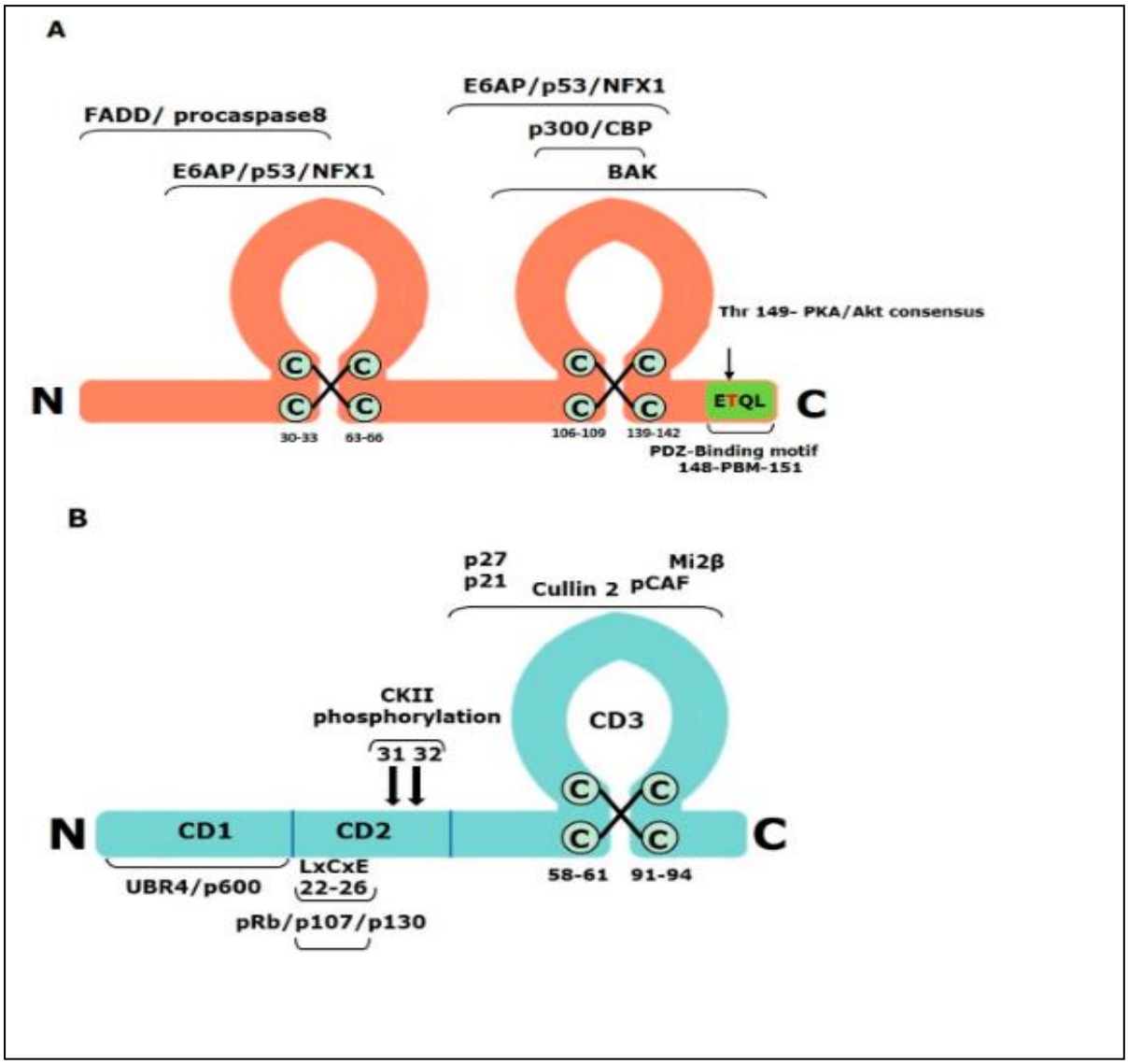


Figure 1.7 (A) Diagrammatic representation of HPV16 E6 showing the position of amino acid motifs that are important for protein integrity and function. (B) Diagrammatic representation of E7 and the most important amino acid motifs required for integrity and protein functions.

(A) The two Zinc finger binding domains are shown, and regions that are involved in interacting with some of its cellular target proteins. The PDZ Binding Motif (PBM) is shown and the PKA phosphorylation site is also indicated; (B) One Zinc finger binding domain is shown, and the pRB binding (LXCXE) and the two serine residues (31 and 32) that are susceptible to casein kinase II (CKII) phosphorylation. Adapted with modifications from (Tomaić, 2016).

The HPV E4 ORF is within the E2 ORF, at about the central portion of the E2 gene within the flexible hinge region encoding domain (Doorbar, 2013). The main E4 gene product, E1^E4, is a spliced mRNA with a start codon in the E1 ORF (donor) and the acceptor site in the E2 ORF (Doorbar, 2013). One of the key functions of E4 in mucosal infections is its ability to restructure cytokeratin filaments in the later stages of the cycle, disrupting their structure and facilitating virion release from host cells (Wang et al., 2004, Doorbar et al., 1991). In cutaneous infections caused by HPV1 and HPV63 (Mu genus), HPV4 and 65 (Gamma genus) and HPV5 and 8 (Beta genus), the E4 protein is responsible for the formation of cytoplasmic inclusion granules that are characteristic of these infections (EGAWA, 1994, Peh et al., 2002). The great abundance of E4 in cutaneous lesions is linked to these inclusion bodies (Doorbar et al., 1996), and have been found mostly in cells supporting vegetative viral genome amplification (Roberts et al., 2003).

L1 and L2 activities during late phase of viral replication: The late phase of the viral replication involves increased production of viral copy numbers of copies per cell. In the upper epithelial layers, viral gene expression shifts to structural proteins, the L1 and L2, in order to promote virion assembly (Cerqueira and Schiller, 2017). The life cycle of PVs is heavily linked to differentiation of epithelial cells to which they are obligate pathogens, thus making the investigation of many aspects of PV biology difficult to study (Cerqueira and Schiller, 2017) [Figure 1.8].

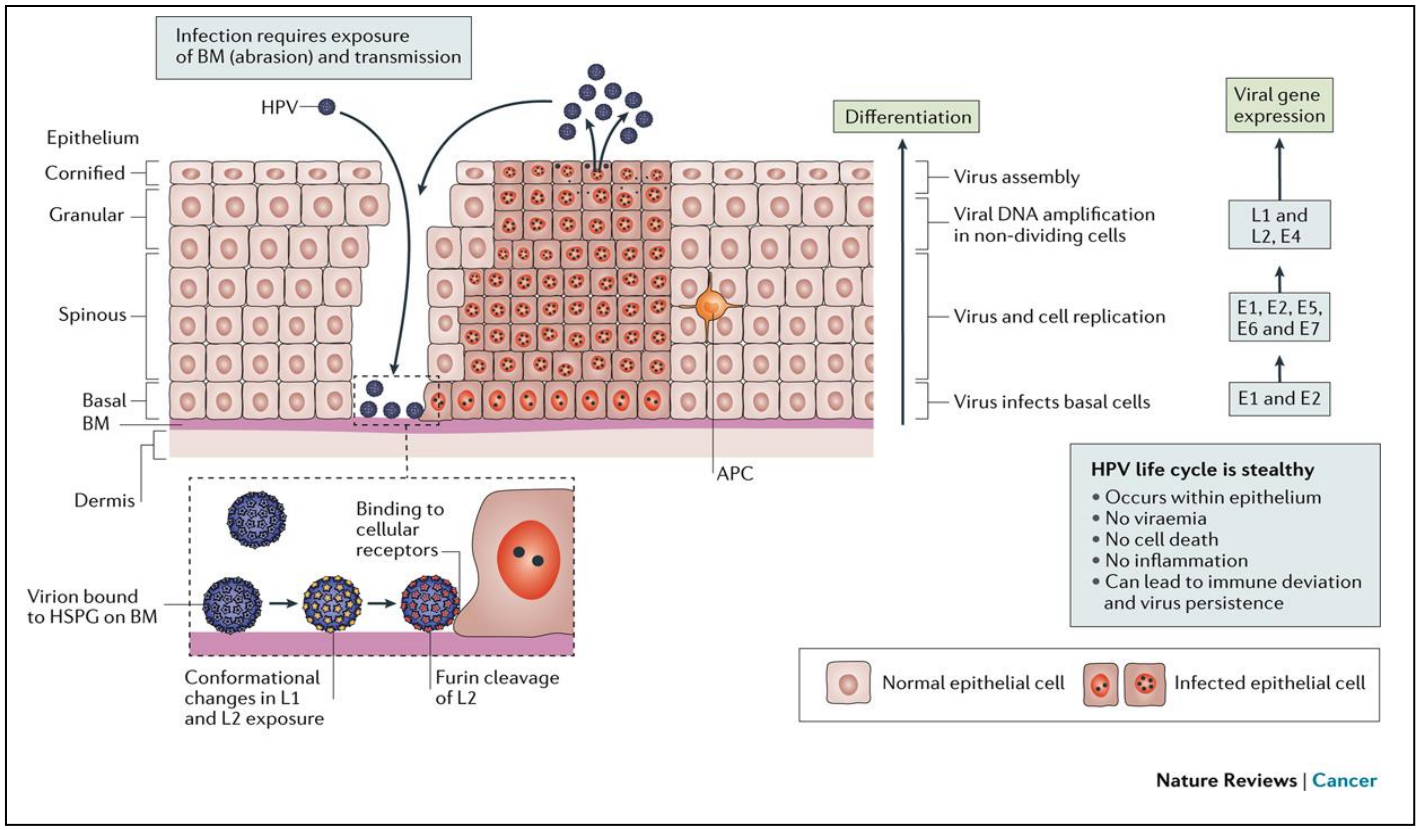


Figure 1.8 HPV Life Cycle and Replication.

Taken from (Roden and Stern, 2018). HPV infection involves the binding of the virus to heparin sulfate proteoglycans (HSPGs) on the basement membrane (BM) through the major capsid protein L1. This is ensued by virus entry into the target basal keratinocyte. Several entry pathways have been proposed, none of which are necessarily independent of the other. In the infected basal cells, the viral genome replicates and establishes ~50 HPV copies, which then segregate between the daughter progeny as the cells undergo cell division. The early viral proteins, E6 and E7, are essential in stimulating the continued growth and environment for E1 and E2-driven vegetative viral genome replication to a very high copy number. Differentiation of infected cells in the upper epithelial layers stimulates the expression of E4 and then L1 and L2 to package the high copy numbers of the viral genome. The virions are released as E4 disintegrates the cytokeratin filaments, and the keratinocyte remains are removed from the epithelial surface. The viral life cycle is accomplished with minimal cell death and without peripheral viraemia or apparent inflammation, and thus evading local immune responses.

The L2 ORF contains early polyadenylation sites, which are positioned in the L2 ORF at the 5' end or the N-terminal, are adenine rich and their function is to process early viral mRNA transcripts see Table 1.1. The late polyadenylation sites are usually downstream of the L1 ORF within the LCR. Polyadenylation sites usually have this signature -AATAAA- (Chen et al., 2007a, Chen et al., 2007b, Ure and Forslund, 2014a, Bottalico et al., 2012, Ostrbenk et al., 2015). The N-terminal of the L2 protein contains a conserved furin cleavage motif (R-X-K/R-R) see Table 1.1. Cleavage of the L2 protein by cellular protease results in viral capsid structural changes resulting in translocation of the viral particles into the host epithelia to deliver and integrate viral DNA into host genome (Ostrbenk et al., 2015). It has recently been described that a transmembrane domain at the N-terminal of the L2 minor capsid protein is essential for translocation of viral DNA across phospholipid bilayer membranes (Bronnimann et al., 2013, Wang and Roden, 2013) see Table 1.1. The domain consists of GxxxG motifs and similar glycine zipper GxxxGxxxG motifs that together work in unison to facilitate packing of DNA helices to pass lipid bilayers. The same NLS signal motifs in E1 and E2 have also been identified in the C-terminus of the L1 and L2 proteins as described elsewhere (Zhou et al., 1991, Nelson et al., 2000) see Table 1.1.

Functions of the LCR during life cycle and replication: The LCR, with no ORFs, encompasses roughly 10% of the genome of most PVs (Bernard, 2013a). The LCR of different HPV types contains varying numbers of palindromic E2 binding sites (ACC-N6-GGT) (Brancaccio et al., 2017) see Table 1.1**.** A palindromic sequence is a nucleotide sequence on a double-stranded DNA where the 5' to 3' forward reading frame on one strand is identical to the sequence reading backward 5' to 3' on the complementary strand. The LCR also has a TATA-

binding protein domain (Brancaccio et al., 2017). The TATA-binding protein (TBP) is a transcription factor that binds to a DNA sequence called the TATA box see Table 1.1, and is important in pre-transcription during viral replication (Bernard, 2013a).

The LCR is involved in regulation of early gene transcription through its interaction with cellular transcriptional factors (Sichero et al., 2013). Some of these transcriptional factors include NF1, API, Oct-1, Tef-1, Tef-2, YY1, TFIID and C/EBP.

NF-1 sites: The NF1 plays a role in transcription during carcinogenesis. It was originally identified in the adenovirus genome where it binds to TTGGC$(N)_6$CCAA but in PVs it binds to the half-site non-palindromic TTGGC (Bernard, 2013b).

AP-1: The AP-1 site has been widely studied, its principal function is in viral gene expression and it has been suggested as the principal activator of all activators. The typical AP-1 binding site is the TGANTCA sequence. Major variations from this sequence have been known to also bind T (G/T)A/T)NT(A/C)A, making it impossible to predict AP-1 binding sites by mere sequence inspection, functional assays are a necessity (Bernard, 2013b).

Oct-1: Octamer-1 is a member of the POU (Pit, Oct and Unc). Oct-1 has been shown to regulate a number of viral and cellular genes. The POU family possess a similar bipartite DNA binding sequence. The Oct -1 binds specifically to an ATGCAAAT sequence. It has also been shown to bind very divergent factors making it difficult to use sequence inspection to identify these binding sites (Tantin, 2013, Bernard, 2013b).

1.4 *PAPILLOMAVIRUS* CONSERVED DOMAINS AND MOTIFS

According to the NCBI Conserved domains database (CDD) definition, conserved domains are evolutionary distinct functional and/or structural units of a protein (Fong and Marchler-Bauer, 2008). Domains are often identified by repeated (sequence or structure) units, which may vary in different contexts. Conserved domains contain maintained sequence patterns or motifs, which enable their identification in polypeptide sequences. Table 1.1 contains a description of each HPV protein and the conserved domains found in them.

We report in the third chapter of this thesis on the different HPV conserved domains that were found in the ten novel Gamma-HPV types. However, the functions of these observed domains are beyond the scope of this thesis as the focus was on the identification and recognition of known conserved sequences.

Table 1. 1 PVs Conserved domains and motif sequence patterns.

Gene/ Protein	Domain	Motif consensus sequences	Reference
E6	GATA Type Zinc fingers	$CxxC(x)_{29}CxxC$	(Ostrbenk et al., 2015)
	PDZ binding domain	x(T/S)x(L/V)	(James and Roberts, 2016, Miller et al., 2012, Songyang et al., 1997)
E7	E7 Conserved region 1		(Wise-Draper and Wells, 2008)
	GATA Type Zinc fingers	$CxxC(x)_{29}CxxC$	(Ostrbenk et al., 2015)
	pRB binding domain	LxCxE	(McLaughlin-Drubin and Munger, 2009a, Mitsuishi et al., 2013, Ure and Forslund, 2014a)
E1	ATP binding site	$G(X)_4GK(T/S)$	(Titolo et al., 1999)
	Bipartite nuclear localisation signal	KRK and KRRL	(Bernhofer et al., 2018, Zhou et al., 1991, Nelson et al., 2000, Ostrbenk et al., 2015)
	Nuclear export domain putative	$(L/I)(x)_{2-3}(L/I)xx(L/I/V)x(L/I/V)$	(Fischer et al., 1995, Bergvall et al., 2013b)
E2	DNA recognition helix	$GxxNxLKCxRxR(x)_8$	(Graham, 2016)
	Nuclear localisation domain	RKRxR/KRRR/KRXR	(Bernhofer et al., 2018, Zhou et al., 1991, Nelson et al., 2000, Zou et al., 2000)
L1	Nuclear localisation like domain	K(K/R)R(K/R)	(Bernhofer et al., 2018, Zhou et al., 1991, Nelson et al., 2000)
L2	Nuclear localisation like domain	$(K/R)_3R(K/R)$	(Bernhofer et al., 2018, Zhou et al., 1991, Nelson et al., 2000)
	Transmembrane binding domain	$G(x)_3G(x)_3G$	(Bronnimann et al., 2013, Wang and Roden, 2013)
	Furin cleavage site	Rx(K/R)R	(Ostrbenk et al., 2015)
	Early polyadenylation site	$AAT(A)_3$	(Chen et al., 2007a, Chen et al., 2007b, Ure and Forslund, 2014a, Bottalico et al., 2012, Ostrbenk et al., 2015)
LCR	E2 binding sites	$ACC(N)_6GGT$	(Androphy et al., 1987, Hirochika et al., 1988, Li et al., 1989, Võsa et al., 2012)
	TATA binding box	$TAT(A)_3$	(Bernard, 2013a)
	Late polyadenylation site	$AAT(A)_3$	(Chen et al., 2007a, Chen et al., 2007b, Ure and Forslund, 2014a, Bottalico et al., 2012, Ostrbenk et al., 2015)

1.5 EVOLUTIONARY DYNAMICS OF *PAPILLOMAVIRUSES*

PVs are thought to evolve slowly because they hijack host cellular high-fidelity DNA polymerases with proofreading activity for replication (with an error rate of about 4.3 x 10^{-5} substitutions per year (Korona et al., 2011)). Hence, it is generally assumed that PVs have co-evolved with their hosts (Bravo and Felez-Sanchez, 2015, Chen et al., 2009, Dube Mandishora et al., 2018). However, several host factors may affect PV evolutionary rates over time; for example the cellular protein APOBEC3 cytidine deaminase may result in codon usage biases in C>>T substitutions, selection pressures due to cellular or humoral immune responses may differ between genes and result in these genes displaying different substitution rates, or the cellular polymerases of different host species may differ in their degree of fidelity and proofreading activity such that virus lineages infecting different hosts might display different substitution rates (de Oliveira et al., 2015). Different PV genes have been shown to have different evolutionary rates, from an overall rate of 1.95 x10^{-8} and a range of 1.76 x10^{-8} to 2.69 x 10^{-8} substitutions/site/year (Van Doorslaer, 2013). The PV genome has modular nature, which is evident today in the different evolutionary rates of the different genes, the new genes E5, E6 and E7 encoding oncoproteins diverge faster than the old four genes E1, E2, L2 and L1 (García-Vallvé et al., 2005). The four old proteins are by themselves able to fulfil the basic tasks of replication, regulation, stabilisation and viral DNA packaging leading to vegetative release of viral progeny from host cells (Longworth and Laimins, 2004). The acquisition of the new oncogenes has introduced two PV phylogenies, high risk Alpha-PVs cluster together according to the phylogeny of these oncogenes (E5, E6, E7) but they do not cluster together according to the phylogeny of the capsid proteins (L1 and L2) (Bravo and Alonso, 2004b). Additionally, the number of mutations are higher in the

new oncogenes than in most of the PV genes (Bravo and Alonso, 2004b). It is thus proposed that the history of PVs took place in different stages, the first stage represents the initial appearance of a prototype-PV with the cardinal genes (E1, E2, L2 and L1) found in all PVs. Then second stage involved further divergent evolutionary processes that led to the acquisition of the E5, E6 and E7 oncoproteins, and these newly acquired proteins evolved about twice faster than the core region of the genome (Garcia-Vallve et al., 2005). This is proposed to have occurred about 150 million years ago. The rapid diversification of the PVs is obvious from the star-like appearance of the PV phylogenetic tree (Garcia-Vallve et al., 2005). It has been estimated that PVs appeared about 250-150 million years ago (Smelov et al.) Smelov et al. (2018) in the Mesozoic Age (Age of reptiles/dinosaurs) (Bravo and Felez-Sanchez, 2015) [Figure 1.9], whereas the most recent common ancestor (MRCA) of the Alpha, Beta, Gamma, Mu and Nu HPVs is predicted to have existed around 30-50 MYA in the Cenozoic Age (Age of mammals) (Chen et al., 2007b) (refer to Figure 1.10). Further, the MRCA of the present-day Gamma-HPVs is predicted to have existed about 15-30 MYA (Chen et al., 2007b). The recent determination of new distinct PVs using next generation sequencing methods has begun to shed light on the evolutionary history of this virus family. However, the evolutionary history is complicated by both inter-gene phylogenetic incongruence and intra-gene variations (Van Doorslaer, 2013).

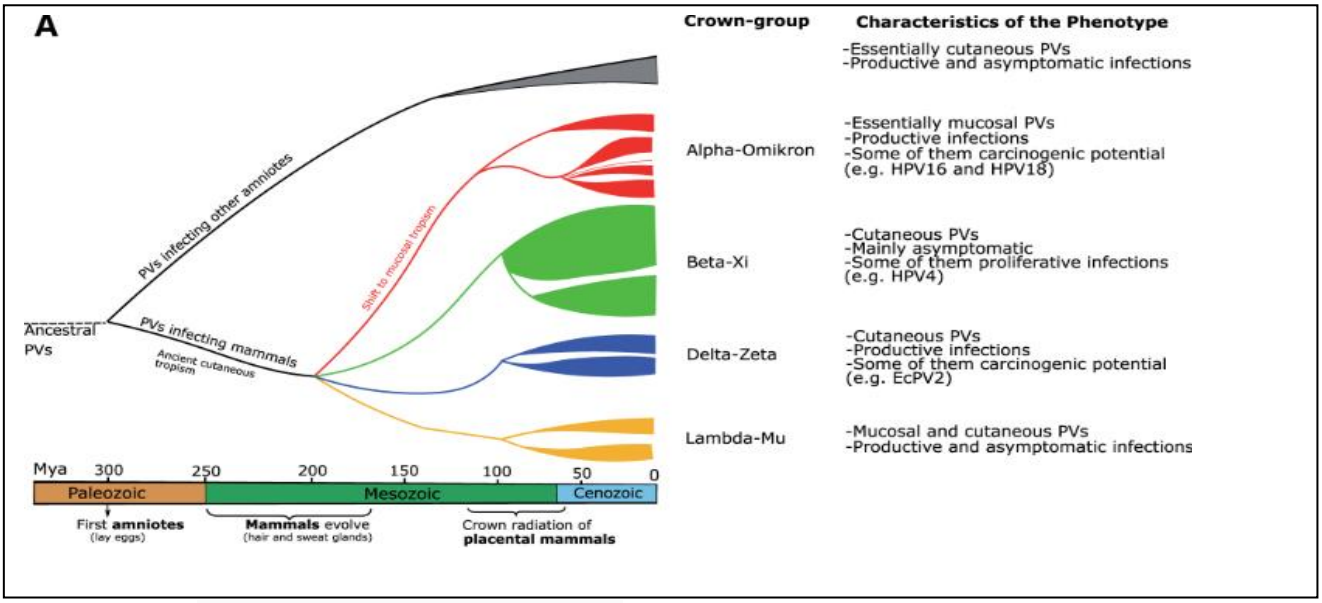


Figure 1.9 Global scenario of PV evolution.

A) Ancestral tetrapod vertebrates were already infected by ancestral PVs. The four PV main groups (labelled in red, green, blue and orange) appeared during the evolution of sweat glands and hairs (250–150 MYA). Ensuing mammalian diversification triggered a further second phase of PV diversification (110-60 MYA) (Bravo and Felez-Sanchez, 2015).

It has been observed that the genes and encoded amino acid sequences of viral early and late proteins have evolved differently in terms of evolutionary rate and selection pressure (Harari et al., 2014, García-Vallvé et al., 2005), and hence the incongruence between early and late trees. PV recombination events may provide a clue to the phylogenetic incongruence, but recombination is not the only explanation for this incongruence. The study of PV recombination has mainly been hampered by limited recombination detection methods, mainly due to technical difficulties associated with the accurate alignment of highly diverse PV gene sequences prior to analysis, misaligned regions of sequences can be mistakenly be identified as recombination origins (Bolatti et al., 2016).

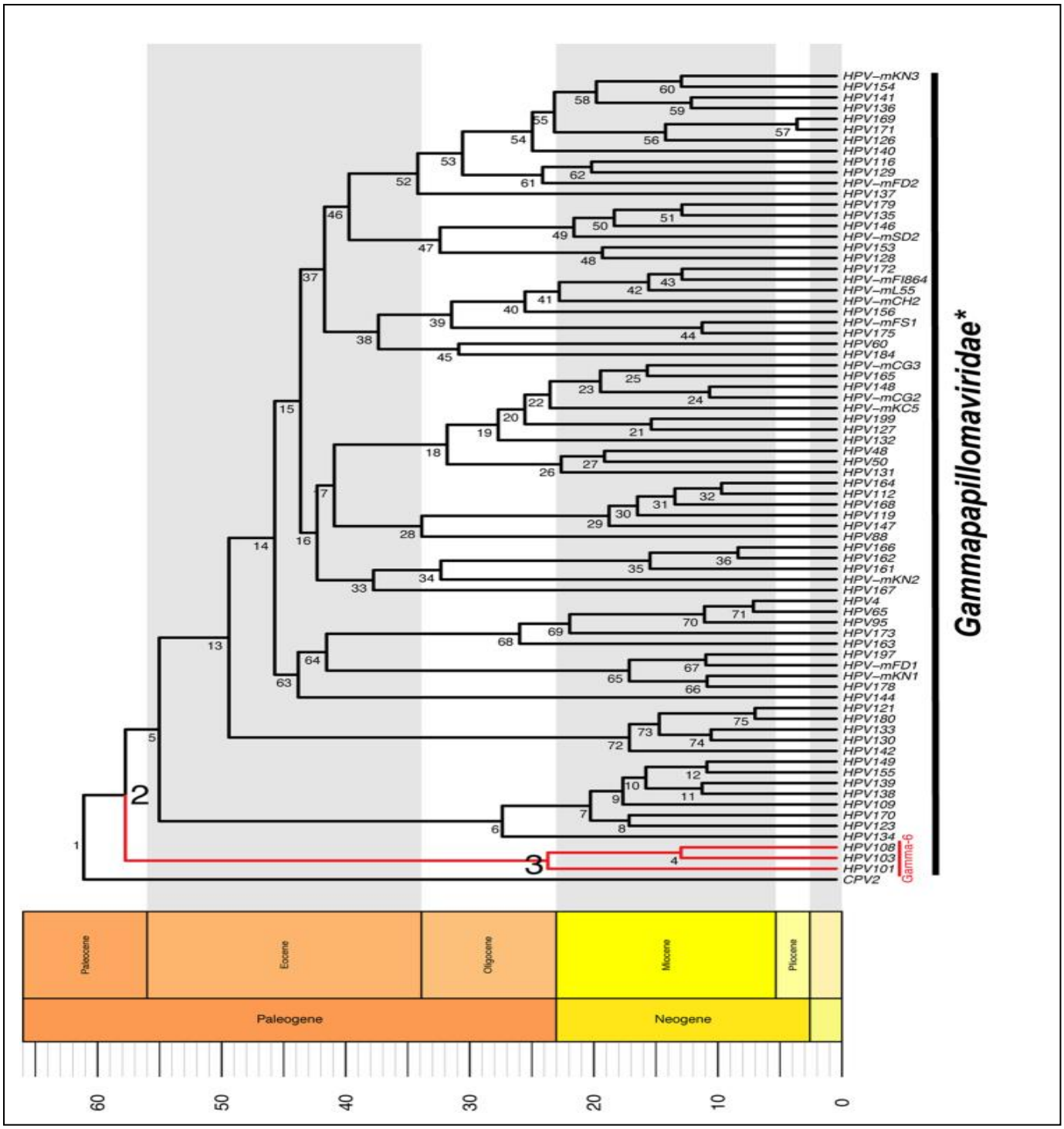


Figure 1. 10 Time calibrated phylogenetic tree of Gamma-HPVs.

Classification based on (Bernard et al., 2010). The asterisk (*) indicates that some members of this clade have not yet been officially recognized as members of the Gamma-HPVs genus. The Gamma-6 species is highlighted in red. The scale bar indicates millions of years before the present. The Gamma-HPV genus tree was rooted on CPV2 (Van Doorslaer and McBride, 2016). Note: the Paleogene and Neogene Ages are Sub-ages of the Cenozoic.

The biological plausibility of HPV recombination is occasioned by the plurality of HPVs and the high frequencies of observed HPV co-infections (Angulo and Carvajal-Rodriguez, 2007). Since the emergence of the AIDS pandemic, AIDS patients have been infected with diverse and multiple HPV types. It is expected that this environment facilitates the emergence of HPV recombinants with somewhat new pathogenic characteristics (Varsani et al., 2006). However, there have been very few studies done on PV recombination, due to lack of proper detection methods until the year 2000, when the Recombination detection program (RDP) method was published by Martin and Rybicki (Martin and Rybicki, 2000), and subsequent versions of the same method in 2005 (Martin et al., 2005a, Martin et al., 2005b). Even after the development of the RDP detection method, very few studies on PV recombination have been published, for Alpha-PVs (Narechania et al., 2005b, Ronco et al., 2014), Gamma-PVs (Bolatti et al., 2016, Varsani et al., 2006) and Beta-PVs (Varsani et al., 2006). The Mu and Nu genera have very few PV types to study recombination unless the current efforts to discover novel HPVs using next generation sequencing are stepped up. However, Shah *et al*. have proposed that inter-genus recombination of Alpha, Beta and Gamma-PVs might have occurred in primates (Shah et al., 2010).

The evolutionary scenario of PVs remains unresolved and it will take more novel types to be discovered and an increment in the number of PV sequences before a conclusive understanding of the origin and adaptive radiation of PVs is completely understood. So far it is widely agreed that these viruses are evolutionarily static, diverge slowly, have strict tissue tropism, are host specific and that there are no recombination events (Van Doorslaer, 2013). It is further assumed that a primary adaptive radiation event of PVs led to the emergence of the first main lineages, and a second adaptive radiation of the main lineages and host

expansion led to species specificity. The third radiation occurred with the emergence of the E5 oncoprotein giving rise three viral lineages (Alpha, Mu and Nu) (Willemsen and Bravo, 2018).

We report in the third chapter of this thesis on the investigation of recombination events among ten novel HPV types HPV211-HPV216 and HPV219-HPV222 (Murahwa et al., 2018) aligned together with all currently known members of the Gamma-HPVs genus and the use of phylogenetic tree incongruence tests to verify or disprove recombination. We also report on divergence times of the ten novel Gamma-HPVs from the MRCA and attempt to provide an evolutionary pathway and mechanism. Understanding the mechanism and way PVs evolved will further clarify the correlation between a given genome and the phenotype and eventually the resulting clinical viral infection.

1.6 HPV AND CANCER EPIDEMIOLOGY

HPVs are the most commonly implicated viruses in a number of human malignancies with 4.5% of all cancers worldwide (with an incidence of 630 000 cases per year) being attributable to HPV infection: 8.6% in women and 0.8% in men (de Martel, 2017). HPV-related cancers account for 38,000 cases of which 21,000 are oropharyngeal cancers occurring in more developed countries, and the relative contributions of HPV16/18 and HPV6/11/16/18/31/33/45/52/58 are 73% and 90%, respectively (de Martel, 2017) [Figure 1.11]. Involvement of HPV in cervical, penile, oral, genital, and oropharyngeal cancers and cutaneous lesions such as skin warts, squamous cell carcinomas, and basal cell carcinomas has been extensively documented (Murahwa et al., 2014, Bouwes Bavinck et al., 2010, Giuliano et al., 2008, Munger, 2002). Epidemiological surveys have documented extensively

the association between HPV and non-melanoma skin cancers (Bouwes Bavinck et al., 2010, Zakrzewska et al., 2012, Iannacone et al., 2013). The involvement of HPV in infections and cancer in a population with a high burden of HIV and AIDS is of great public health concern.

Table 1. 2 International Agency for Research on Cancer Risk classification of HPVs

Classification	**HPV types**
Group 1 Carcinogenic (high-risk)	16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59
Group 2A Probably carcinogenic (high-risk)	68
Group 2B Possibly carcinogenic (intermediate risk)	26, 53, 66 [a], 67, 70, 73, 82 **Beta-HPV types implicated in EV disease** **HPV5 and HPV8**
Group 2B Unknown risk [b]	30, 34, 69, 85, 86, 97
Group 3 Low-risk	6, 11, 40, 42, 43, 44, 54, 61, 71, 72, 81, 89, 90 [c] **HPVs of the Beta genus except types 5 and 8**
Not included in IARC 2012 classification (unknown risk/low- risk)	HPV 74, 83, 87, 91

Adapted from (Meisal et al., 2017)
a-Classified as carcinogenic (high-risk) in IARC 2007
b-Classified in this group based on their phylogenetic analogy to HPV types with sufficient or limited evidence of carcinogenicity.
c-Not specified in IARC 2012.

HPVs, that are associated with malignant anogenital cancers, are found in the Alpha-PV genus and the International Agency for Research on Cancer 2007. (2007) defined twelve HPVs as 1A carcinogens and a thirteenth type as a 2A carcinogen (probably carcinogenic) (Meisal et al., 2017). Beta-PV infection is found in EV, a rare inherited disease associated with a high rate of skin cancer with the genus Beta- HPV types 5 and 8 classified by IARC as "possible carcinogenic" biological agents (group 2B) in EV disease (Smola, 2014) [Table 1.2]. A systematic review of case-control studies on cutaneous HPVs showed that several Beta-PV species were detected significantly more often in squamous cell carcinoma than among controls but these Beta-PVs have not yet accepted as the aetiological agent for these cancers (Bzhalava *et al*. 2013). Although Beta-PVs have been regarded as “cutaneous” and not “mucosal” PVs, there is a report of anal Beta-PV infections being highly prevalent in the men who have sex with men (MSM) population and that Beta-PVs can establish persistent infection in the anal region for up to 4.8 years (Mlakar et al., 2014).

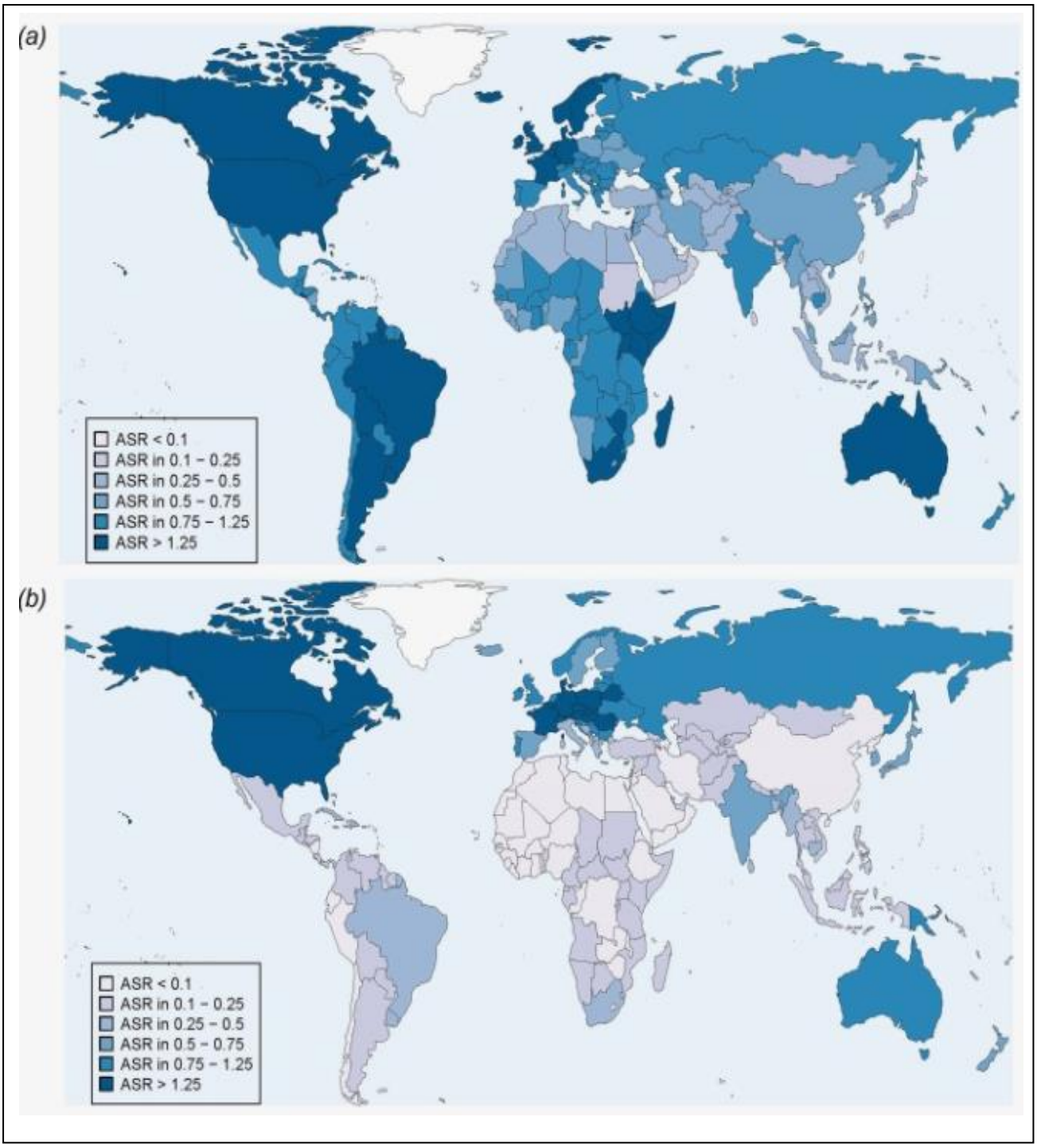


Figure 1.11 Age standardized world incidence rate (per 100 000) cancers linked to HPV, panel (a) anogenital cancers and (b) head and neck cancers (de Martel, 2017).

1.7 GAMMAPAPILLOMAVIRUSES AND CANCER

Among all the HPVs, the genus Gamma-HPVs is the most diverse and fast growing with 27 species and 98 officially recognised genotypes (Mühr et al., 2018) [Figure 1.12]. Gamma-HPVs appear to have broad epithelial tissue tropism with reported detection in cutaneous, mucosal and mucocutaneous sites (Ure and Forslund, 2014b, Ma et al., 2014b), including healthy skin (Antonsson et al., 2000), cutaneous lesions (Ekström et al., 2011), gut (Ma et al., 2014b), penis (Sichero et al., 2013, Sichero et al., 2014b), oral mucosa (Bottalico et al., 2011b), nasal mucosa (Forslund et al., 2013b), anal canal (Sichero et al., 2015b) and cervical mucosa (Meiring et al., 2017) .

These viruses have not been conclusively associated with any pathology or oncogenicity. A recent systematic review (Bzhalava et al., 2013) did however describe significantly elevated antibody levels in squamous cell carcinoma (SCC) cases infected with HPVs from the Gamma-1 species. While Gamma-HPVs remain under-studied at a molecular level, a recent proteomic study of HPV197 demonstrated that the E6 and E7 proteins of this Gamma-HPV interact with several cellular targets including some of the important mediators of the oncogenic activities of high-risk E6 and E7 proteins (Grace and Munger, 2017). The E7 protein of Gamma-6 HPV108 induces abnormal growth in organotypic keratinocyte cultures (Nobre et al., 2009b). The potential oncogenic activity of Gamma E6 and E7 proteins requires further investigation. Interestingly, HPVs in the Gamma-6 species do not encode E6 and E5 proteins (Ameur et al., 2014a, Chen et al., 2007b).

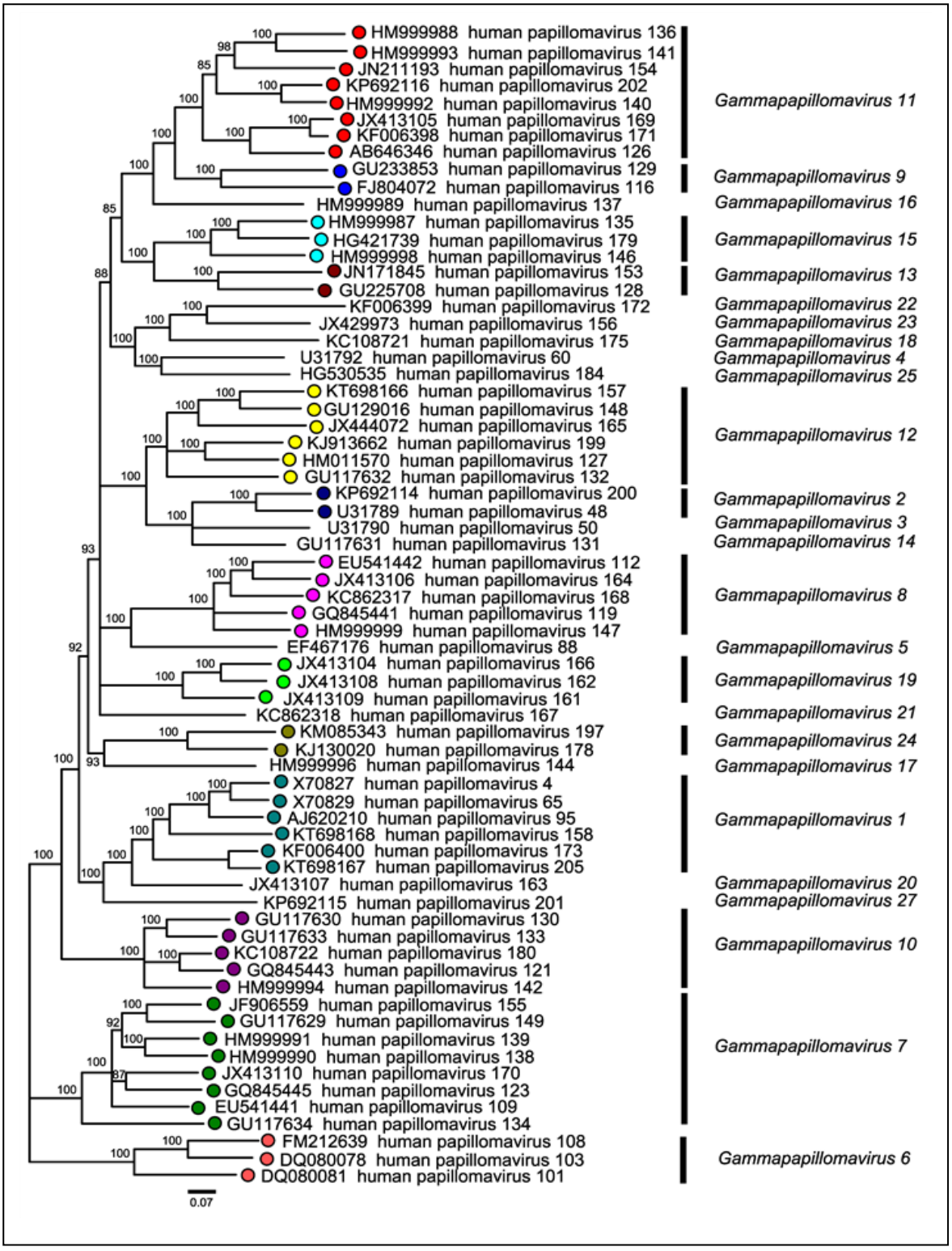


Figure 1.12. Phylogenetic tree of members of the genus Gamma-HPVs.

Taken directly from ICTV website https://talk.ictvonline.org/ accessed 18/09/2018. The E1, E2, L2, and L1 nucleotide sequences of 343 PV isolates including representatives of all species and genera within the *Papillomaviridae* family were aligned as amino acid sequences using MUSCLE v7.221. The JModeltest2 was used to determine the optimal model of evolution (GTR + I + G) for the concatenated nucleotide sequences. Maximum likelihood (Mlakar et al.) trees were constructed using RAxML MPI v8.2.9 implementing the GTR substitution model. ML bootstrap analysis used the autoMRE-based stopping criterion in RAxML. Following tree construction (tree available in the Resources section of the *Papillomaviridae* Report), the subtree corresponding to the genus Gamma-HPVs was isolated. The tips are labelled with virus names and accession numbers; nodes are labelled with bootstrap support values (Simmonds et al., 2017).

A recent study showed a high proportion of Gamma-HPVs in pre-cancerous cutaneous lesions of immunocompetent individuals, giving insights into the possible early role of these viruses in cutaneous malignancies (Bolatti et al., 2018). It was recently shown that Gamma-6 HPVs or rather E6-minus viruses have acquired a 37 amino acid protein named E10, the ORF encoding this protein is upstream of the E7 start codon (Van Doorslaer and McBride, 2016, Van Doorslaer et al., 2017a). However, the acquisition of E10 does not likely compensate for all E6 functions.

HPVs belonging to the Gamma-HPVs have been always classified as cutaneous types (de Villiers, 2013). Recent evidence suggests a wider broader tissue tropism, with the cutaneotropic viruses being detected in muco-cutaneous areas of anogenital region, oral and nasal mucosa and genital lesions (Moscicki et al., 2017, Bottalico et al., 2011c, Foulongne et al., 2012, Ma et al., 2014a, Sichero et al., 2013, Forslund et al., 2013a, Tommasino, 2017). Our preliminary data shows that these viruses as well as Gamma-PVs are common in penile samples (Meiring et al., 2017). There is a paucity of information on genital Beta- and Gamma-PVs in South African men or women, with only a few studies done (Meiring et al., 2017, Ameur et al., 2014b).

1.8 HPV INFECTIONS IN MEN

Globally HPV accounts for more than half of infection related cancers in women and about 5% in men (Grce and Mravak-Stipetić, 2014, zur Hausen, 2009). As a result of this high association between cervical cancer and HPV in women, men have not been primary focus for HPV research and intervention (Dunne et al., 2006). However, although a neglected field the HPV prevalence among men is high and not low (Rodriguez-Alvarez et al., 2018). The risk factors of HPV in men have been associated with sexual promiscuity, related to number of sexual partners (Nielson et al., 2009). Penile circumcision has been reported as a protective factor against HPV infection (Rositch et al., 2012). Firstly, the presence of a foreskin provides a moist environment that is favourable for HPV survival and facilitates virus entry. Second, the keratinisation of the circumcision scar can reduce HPV exposure and access to epidermal basal cells (Rositch et al., 2012, Olesen et al., 2017). Smoking has also been found to associated with HPV infection as cigarette smoke contains chemicals that disrupt the genetic intergrity of human cells which accelerates progression towards cancer in men (Rocha Rodríguez et al., 2012). Condom use is considered as partially protective against HPV infection (Lu et al., 2009, Repp et al., 2012, Vardas et al., 2011). Nielson *et al*. report that consistent condom use is associated with lower HPV prevalence in men (Nielson et al., 2007). Socio-economic factors have also been reported to have an impact in HPV infection among men, and among these are level of education (Lu et al., 2009). College and other forms of higher education were significantly associated with lower risk of acquiring new HPV infections (Lu et al., 2009).

Based on current evidence, there is great benefit in vaccinating men against HPV (King et al., 2015). A study done in South Africa also reported on the benefits of vaccinating children in

order to reduce the burden of HPV related cancers such as anal and penile cancer in men (Muller et al., 2016).

Racial and socio-cultural differences have been seen to be a factor in HPV acquisition, in a study by Akogbe *et al*. (Akogbe et al., 2012) (part of the multinational HIM study) the prevalence of any HPV infection and of oncogenic HPVs was lowest in Asian and Pacific Islanders, and this was attributed to their strict social behaviour associated with a conservative and restricted sexuality. There is limited data in Africa and particularly southern Africa in terms of HPV in men, and very little information on socio-economic and socio-cultural factors in HPV acquisition.

Anal HPV16 infection is known to be increased by the practice of anal sex which is common among men that have sex with men (MSM) and further amplified by HIV associated immunosuppression (Marra et al., 2018). It has been shown that anal cancer incidence is higher in MSM than in men who have sex with women (MSW) (Colon-Lopez et al., 2018, Silverberg et al., 2012), thus establishing HIV infection and sexual preference as independent determinants of anal HPV infection in men. The anogenital distribution and concordance of Beta- and Gamma-PV types in MSW and their partners have also been studied (Smelov et al., 2018). It was shown that there was a high concordance of Beta- and Gamma- -HPVs among sex partners although the etiological role, if any, in carcinogenesis was not clear.

The presence of HPV DNA does not necessarily translate to presence of an infectious virus (Di Bonito et al., 2015), but more puzzling is the unclear transmission route of cutaneous Beta-and Gamma-HPVs into the anogenital area. It has been proposed that the presence of cutaneous HPV DNA in the anogenital area may be reflective of the deposition of viruses

released from other body parts with real infections (Sichero et al., 2013). Most of the recent information on HPV infection in men has been obtained from the HPV in men (HIM) multi-centre study with over 4000 participants, which has described the prevalence of cutaneous HPV types among samples collected from external genital skin (foreskin etc.), anal canal and oral cavity from men (Bottalico et al., 2011a, Forslund et al., 2013c, Sichero et al., 2014a, Sichero et al., 2015a, Torres et al., 2015, Agalliu et al., 2016).

In this thesis, the isolation of novel HPV types from penile swabs, from black South African men that had female sex partners, is described. The diversity of HPV viruses isolated from these swabs was vast as reported by Meiring et al (Meiring et al., 2017).

1.9 MULTIPLE HPV INFECTIONS

It has been reported that certain combinations of HPV types increase the risk of invasive cervical cancer (ICC) and pre-cancerous lesions when compared to infection with the individual HPV types (Carrillo-Garcia et al., 2014, Williamson, 2015). The role of Beta- and Gamma-PVs in this process is unknown. HIV-1 positive women have a higher prevalence of multiple HPV type infections in normal, premalignant lesions and ICC (Denny et al., 2014, Mbulawa et al., 2012). HPV prevalence is very high in HIV-infected people, but generally only mucosal HPV types have been published on. Our preliminary study indicates a higher prevalence of Beta- and Gamma-PVs in HIV-positive versus HIV-negative men from Cape Town (Meiring et al., 2017).

1.10 HPV AND HIV CO-INFECTIONS

The role of multiple HPV infections, including Beta- and Gamma-PVs on susceptibility to HIV is not well understood. There are a couple of studies indicating that people infected with

genital HPVs are at higher risk of acquiring HIV-1 than people who are not infected with HPV, even after adjusting for risk factors (Williamson, 2015). In a meta-analysis HIV-1 incident infection was significantly associated with high-risk HPV infection in 5 of 6 studies (Lissouba et al., 2013). The mechanism for this increased susceptibility to HIV is not well understood. A study on Zimbabwean women showed that incident HIV-1 infection was directly linked to HPV clearance, suggesting that the immune response to HPV may result in an increase in cells prone to HIV infection in the genital tract. HIV-1 acquisition was also linked to HPV clearance in a study on circumcised men, with increased HIV-1 acquisition being associated with number of HPV genotypes cleared (Tobian et al., 2013). The same study showed that increased epidermal dendritic cell density was associated with HPV clearance and increased the risk of HIV-1 acquisition (Tobian et al., 2013). Thus, the immune response to HPV in mucosal tissue may result in an increase in the number of cells that are susceptible to HIV-1 infection, and accordingly result in an increased incidence of HIV.

1.11 HPV GENOMICS IN THE NEXT GENERATION SEQUENCING ERA

Traditionally, the classification and description of novel viruses by the ICTV requires substantial biological information such as replication cycle, virion morphology, serology, host range, pathogenicity, epidemiology and nucleic acid sequence (Simmonds, 2015). However, next generation sequencing (Van Doorslaer et al. (2017a) and metagenomic approaches pose a challenge to the virus classification landscape, with many viruses being known based on sequence data only (Simmonds et al., 2017). As of June 2016, the ICTV approved proposals to classify viruses solely on the basis of metagenomic sequence data, and a consensus statement released in the August of 2016 stated: "*We believe that the time has come to advance the philosophy and practice of virus taxonomy by admitting viruses that are*

identified only from metagenomics data as being bona fide viruses…" (Simmonds et al., 2017). HPV classification based on next generation sequencing (NGS) data alone preceded that of many other virus families. Table 1.3 shows some of the most recent HPVs discovered by NGS methods and assigned HPV reference centre numbering but most are pending ICTV classification. It has been shown that rapid and accurate HPV genotyping for early cancer diagnosis can also be facilitated by NGS (Nilyanimit et al., 2018), as it detects HPV types that are not usually detected by routine genotyping methods.

At the University of Cape Town (UCT), we have experience using next generation sequencing platforms to characterize HPVs from cervical samples and identify viruses that are not detected by commercial HPV typing kits (Ameur et al., 2014b, Meiring et al., 2012). In the one study, 16 HPV genotypes were detected in a selected specimen using illumina sequencing. Four of the genotypes (HPV-30, 74, 86 and 90) detected were not included in the commercial kit. The prevalences of HPV-30, 74, 86 and 90 in 109 HIV positive South African women were found to be 14.6%, 12.8%, 4.6% and 8.3% respectively illustrating that these are not rare viruses in this population (Meiring et al., 2012). In the second study, a total of 46 different HPV types were found, many of which are not detected by existing genotyping assays and two novel HPVs identified by NGS in Cape Town were found to lack the E6 gene found in the majority of described HPVs (Ameur et al., 2014b).

Table 1. 3 Recent novel HPVs discovered by Next generation sequencing.

Author/dateNovel HPV type		Complete genome amplification-Cloning system	Next Generation Sequencing Facility/Bioinformatics
Hosnjak L et al. 2015 (Hosnjak et al., 2015)	179 and 184	Two overlapping amplicons-TOPO XL PCR	Microsynth AG facility (Swiss) + primer walking/Vector NTI Advance-MEGA5
Chouhy D et al. 2013 (Chouhy et al., 2013)	156	Two overlapping amplicons-pGEM T easy	University of Maine facility (Araldi et al.)/MEGA5-BEAST
Mtsuishi T et al .2013 (Mitsuishi et al., 2013)	160	Single 8kb fragment-pBulescript II SK.	Primer walking on AB/GENETYX-ClustalW-Bioedit-MEGA
Bottalico D et al. 2012 (Bottalico et al., 2012)	120	Two overlapping fragments-TOPO TA	Microsynth AG facility (Swiss/MEGA5-MUSCLE-RAxML
Chen Z et al.2007 (Chen et al., 2007a)	101 and 103	Two overlapping fragments-pGEM T easy	Einstein facility/ClustalX-Codon align-Mrbayes
Ure AE et al. 2014 (Ure and Forslund, 2014b)	154	Four overlapping amplicons-TOPO TA	MWG-Germany/UGENE-SMART-MUSCLE
Arroyo S et al. 2014 (Arroyo Muhr et al., 2014)	197	Three overlapping fragments-TOPO PCR kit	Miseq & Sanger/SOAPdenovo-Trinity-IDBA-UD assemblers
Chen Z et al. 2007 (Chen et al., 2007a)	102 and 106	102-two overlapping fragments-pGEM T easy 106 three overlapping fragments.	Einstein facility + primer walking/ClustalX-Codon align-Mrbayes
Kovanda A et al. 201 1(Kovanda et al., 2011)	125	Single PCR fragment-reverse long-range PCR-Clone JET PCR Cloning	Primer walking by ABI Prism/MEGA4-Bioedit
Kocjan JB et al. 2013 (Kocjan et al., 2013b)	174	Single 8kb fragment-TOPO XL PCR	Microsynth AG facility (Swiss)/Vector NTI Advance 11.
Bolatti M et al. 2017 (Bolatti et al., 2017)	209	Two overlapping fragments-TOPO TA and pGEM T easy	University of Maine DNA sequencing facility
Dutta S et al. 2017 (Dutta et al., 2017)	217 and 218	Single 8kb fragment -TOPO XL PCR cloning kit	GATC Biotech, Germany
Brancaccio et al. 2017 (Brancaccio et al., 2017a)	224	Single 8kb fragment -TOPO XL PCR cloning kit	GATC Biotech, Germany
Kocjan JB et al. 2013 (Kocjan et al., 2013a)	159	Two overlapping fragments-TOPO XL PCR	Microsynth AG facility (Swiss)Vector NTI Advance-MEGA5
Latsuzbaia A et a.2018(Latsuzbaia et al., 2018)	226	Single 8kb fragment -TOPO XL PCR cloning kit	Illumina Miniseq platform
THIS STUDY (Murahwa et al., 2018)	211 , 212, 213, 214, 215, 216, 219, 220, 221, 222	Single 8kb fragment-pGEM T easy and TOPO XL PCR	Illumina at Macrogen facility/ CLC main workbench

1.12 BACKGROUND DATA FOR THIS STUDY

There are now more than 221 HPV types described with only the Alpha-PVs and a few Beta-PVs being associated with cancers – the other two major genera infecting humans are the Beta- and Gamma-HPVs (Arbyn et al., 2014) and very little is known about most of these types and their role in genital disease. There are a number of publications studying Beta- and Gamma-HPVs in various skin cancers and pre-cancers but no firm association has been found because of the large variety of viruses detected (Ekstrom et al., 2013a). Other studies have reported a large diversity of Beta-PVs in penile samples (Sichero et al., 2013), including penile carcinomas (del Pino et al., 2012) and genital warts (Bottalico et al., 2012).

Our laboratory at UCT has stored genital samples collected from heterosexually active couples (Mbulawa et al., 2010). These samples have been well characterised for mucosal HPV types detected by the Roche HPV Genotyping linear array test which detects 37 HPV types. A study was done on 218 penile samples (104 HIV negative and 114 HIV positive) using high throughput sequencing (Roche 454) of amplimers (approximately 480 bp) obtained using FAP59/64 primers which have been designed to detect "cutaneous" or Beta- and Gamma-PVs (Meiring et al., 2017) . A total of 181 different HPVs were detected in 1330 incidents of infection. A total of 45 known Alpha-HPV types, 45 Beta-HPV types (34 known, 10 putative and 1 novel putative), and 91 Gamma-HPV types (26 known, 51 putative and 14 putative novel) were detected. Many HPV types (34.3%; 62/181) occurred only once in the penile samples, the majority of these belonged to the Gamma genus (44, 70.9%), followed by 16 (25.8%) Beta-HPVs and 2 (3.2%) Alpha-HPV types. The most frequently detected Gamma-HPVs were HPV121 (6.9%), HPV101 (4.1%) and a novel putative HPV type, CT06 (4.1%). In HIV-positive men types HPV121 (7.9%), HPV101 (6.1%) and CT02 (6.1%) and in

HIV-negative men types HPV121 (5.8%), HPV144 (4.8%), CT06 (3.8%) and CT04 (3.8%) were the most prevalent types. Of the 44 Gamma-HPVs detected in more than one sample, 28 had higher prevalence in HIV-infected men. In total, there were 15 putative novel HPV types found with a prevalence varying between 0.5% to 4.1% of the men sampled (Meiring et al., 2017). The putative novel types L1 short fragments were deposited in Genbank under the following accession numbers (KY062999-KY063013) and were defined as those that had <90% identity to any known or previously described putative HPV.

In this thesis, we report on the discovery, isolation, cloning, whole genome sequencing, characterisation, phylogeny and lastly the evolutionary dynamics inferred from incongruence, recombination and phylodynamic analysis (molecular divergence) of ten of the fifteen novel HPVs.

1.13 JUSTIFICATION

As more and more novel HPV types continue to be discovered worldwide, there is an increasing need to revisit archives of specimens stored and previously tested by first generation of molecular techniques to ascertain the full range of HPV types. Classification and assignment of novel HPVs to types or species can only be correctly achieved by whole genome sequencing. Thus, next generation sequencing technologies provide a powerful tool in attaining this goal. Ascertaining the new HPV types also provide essential information on the development of epitopes for future vaccines since full sequences for development of recombinant proteins will be known.

Very little is known concerning Gamma-PVs in genital infections and there is paucity of information about these viruses in Africa. This thesis focuses on the (i) genomics, (ii) intra-

sample variation and (ii) evolutionary dynamics of ten novel Gamma-HPVs isolated from penile swabs in South Africa.

1.14 AIMS AND OBJECTIVES

1. The characterisation of potential novel HPVs from penile samples by complete genome sequencing and comparison to known HPVs.
2. To identify and describe the conserved functional domains in the proteins of the novel HPV types.
3. To study the phylogeny and genomic variation of the novel HPV types within samples.
4. To explore the evolutionary dynamics of Gamma-PVs by analysing phylogenetic incongruence, recombination and phylodynamics (molecular divergence) in these viruses.

Chapter 2: Discovery, Characterisation and Genomic Variation of Ten Novel *Gammapapillomavirus* Types Isolated from Penile Swabs.

2.0 INTRODUCTION

In this chapter, we report on the discovery, genomic characterisation and phylogenetic evaluation of ten novel Gamma-HPV types: HPV211, HPV212, HPV213, HPV214, HPV215, HPV216, HPV219, HPV220, HPV221 and HPV222. These HPVs were previously identified in a study that was done on 218 penile samples (104 HIV negative and 114 HIV positive) using high throughput sequencing (Roche 454) of amplimers obtained using FAP59/64 primers which were designed to detect "cutaneous" or Beta- and Gamma-HPVs (Forslund et al., 1999). Fifteen putative novel HPV types were identified from the short HPV L1 FAP fragments HPV211 (CT02, KY063000), HPV212 (CT03, KY063001), HPV213 (CT04, KY063002), HPV214 (CT06, KY063004), HPV215 (CT07, KY063005), HPV216 (CT12, KY063010), HPV219 (CT01, KY062999), HPV220 (CT08, KY063006), HPV221 (CT09, KY063007) and HPV222 (CT15 KY063013) with a prevalence varying between 0.5% and 4.1% of men sampled (Meiring et al., 2017). We further examined variation of the novel types in clinical specimens from which they were identified.

2.1 METHODS AND MATERIALS

2.1.1 ETHICS STATEMENT

Ethical approval for the study was granted by the Health Research Ethics Committee of UCT, Faculty of Health Sciences (HREC reference: 231/2015 and 258/2006). Written consent was obtained from all the study participants, after they consented to have their blood and genital swabs samples collected and the storage of their biological samples for future laboratory tests.

2.1.2 PENILE SAMPLES

In 2007, samples were obtained from African black men who were participating in a heterosexual couples study on the natural history of HPV infection in South Africa (Mbulawa et al., 2009). Penile swab samples were obtained by dry swabbing of the penile shaft, glans and foreskin if present, using a Digene swab and stored at -80 °C in specimen transport medium (STM, Qiagen). The details of the original source participants from which the novel HPV types were isolated are given in Table 2.1.

Table 2.1 Details of participants from which novel sequences were obtained.

Sequence ID	Isolate	Age	Isolation source
HPV211	CT02	29	Penile swab from a South African HIV negative male
HPV212	CT03	42	Penile swab from a South African HIV positive male
HPV213	CT04	28	Penile swab from a South African HIV negative male
HPV214	CT06	24	Penile swab from a South African HIV negative male
HPV215	CT07	41	Penile swab from a South African HIV negative male
HPV216	CT12	45	Penile swab from a South African HIV positive male
HPV219	CT01	28	Penile swab from a South African HIV positive male
HPV220	CT08	31	Penile swab from a South African HIV positive male
HPV221	CT09	23	Penile swab from a South African HIV positive male
HPV222	CT15	51	Penile swab from a South African HIV negative male

2.1.3 NUCLEIC ACID ISOLATION AND AMPLIFICATION

DNA was extracted using the MagNA Pure Compact Nucleic Acid Isolation kit (Roche, USA) and circular genomes enriched using the Illustra TempliPhi 100 Amplification kit (GE Healthcare, Amersham, UK). Rolling circle amplification was performed according to the manufacturer's instructions, with initial denaturation at 95^{o}C for 3 minutes and held at 4^{o}C for 10 minutes before adding the buffer enzyme mix, followed by an 18-hour incubation at 30^{o}C and the reaction stopped by a 65^{o}C incubation for 10 minutes.

2.1.4 COMPLETE GENOME AMPLIFICATION

The complete genomes were amplified by PCR using primers (primer sequences are provided in Table 2.2), that annealed back to back allowing for the amplification of the whole genome as one amplicon. Primers were designed based on L1 FAP sequences from Meiring and co-workers (Meiring et al., 2017). Primer sets were designed in CLC main workbench (Qiagen). The partial L1 nucleotide sequences of HPV211, HPV212, HPV213, HPV214, HPV215, HPV216, HPV219, HPV220, HPV221 and HPV222 were originally identified as potentially novel in 2014 using FAP primers on penile samples from men enrolled in a cohort of heterosexual couple.

Table 2.2 Novel HPV L1 back to back type specific primer sequences.

HPV type	Primer Sequences
HPV211	forward-GTTACGGGGAATTCAGATAGGTAGAGGTGG
	reverse- TTCCAAACCAGTCTTTCATGGTCAGAATTG
HPV212	forward- ATAGAAATAGGTAGAGGTGGGCCTTTAGG
	reverse-AGCTCGTAATTTCCAAACTAAGCGTTCG
HPV213	forward-TGGCAGTTACGGGGTGTTGAGGTAGAC
	reverse- CACAAGACGTTCCTCGTTGGGATCATAC
HPV214	forward-GCCTTTGGGTATTGGGTCTACTGGTCAC
	reverse-CCACCCCGTGCAATATCAATGCCACGTAAC
HPV215	forward- ATGGTTTGCAAATTGACAGAGGTGGTCC
	reverse- ATAATTTCCACACCAGCCTTTCATGTTGTG
HPV216	forward- CAACAGGTCATCCATTATTTGATCGCTTAC
	reverse- TTCCTATCCCTAGGGGACCACCTCTATC
HPV219	Forward- TAAGTTATCAGATACCGAAAACCCTTTGGCTC
	Reverse- TTTAATAAAGGATGTCCTGTAGCGCCAATACCC
HPV220	Forward- ATTGAGGTTGGCCGTGGTGGTC
	Reverse- GCCCCGTAGCCTCCATATTAAACGTTCG
HPV221	Forward- GTATAGAAATTCAAAGAGGGGGCCCATTAG
	Reverse- CTTTTAGTTTCCATACCAAGCGTTGGGTTTCTG
HPV222	Forward- CAACAAAATGGCCTTTGTCCTCC
	Reverse- TTGTTCCCCTGCACAAGGTAATGC

The KAPA LongRange HotStart PCR kit (KAPA Biosystems, USA) was used to amplify the novel HPV genomes in a touch down PCR. Each reaction had 1X KAPA LongRange Buffer,

1.75 mM $MgCl_2$, 0.3 mM of each dNTP, 0.675 U of the KAPA LongRange DNA polymerase, 0.5 µM of each primer and nuclease free water. The PCR conditions were as follows: 94°C for 3 min followed by 12 cycles of 25 sec at 94°C, 68°C (reduced by 2°C every 2 cycles) for 15 sec, and 7 min of extension at 68°C, followed by 20 cycles of 25 sec at 94°C, 56 °C for 15 sec and 7 min at 68 °C. The final extension was performed at 72°C for 10 min. The PCR products were purified by gel electrophoresis before proceeding to cloning. For each of the novel HPV types, 2 sets of PCR reactions were done, in order to rule out PCR artefacts in the analysis of sequences.

2.1.5 CLONING

The resulting PCR products were gel purified using the MinElute gel extraction kit (Qiagen, Germany) and cloned into the pGEM-T Easy vector (Promega, USA) for HPV211-HPV216 and the TOPO XL vector (Thermo Fisher Scientific, USA) for HPV219-HPV222. Transformation was done in JM 109 *E. coli* chemically competent cells (Promega, USA). The cells were cultured on Luria broth agar plates containing 50 µg/ml of carbenicillin and incubated at 37°C for 18 hours. Bacterial colonies with recombinant plasmids were selected by blue/white screening method using 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-gal) treated LB agar plates. Ten colonies (ten clones) were selected to represent each of HPV211, HPV212, HPV213, HPV214, HPV215, HPV216, HPV219, HPV220, HPV221 and HPV222 (multiple separate clones of each type were generated in case errors had occurred during PCR amplification). The plasmid DNA of each of the ten clones was isolated from bacterial culture in 3 ml of LB broth containing carbenicillin using High Pure Plasmid Isolation kit (Roche, USA). Aliquots of the purified plasmid DNA were analysed by gel

electrophoresis and compared to a known pGEM-T Easy vector or TOPO XL Vector with HPV DNA fragment for confirmation.

Successful cloning was also confirmed by Sanger sequencing of approximately 800 bp of the HPV plasmids at the insert and vector junction using an M13 reverse primer. Subsequent alignment of the sequenced fragments to the NCBI non-redundant nucleotide database using BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was done to verify that they were HPV sequences.

2.1.6 LIBRARY PREPARATION AND SEQUENCING

The purified plasmids were then prepared for next generation sequencing (Van Doorslaer et al.). Libraries were prepared using KAPA HyperPlus library preparation kit (KAPA Biosystems) which involved enzymatic fragmentation, end repair and A tailing, library amplification and a last clean up step using solid phase reversible immobilisation (SPRI) were done as described below.

DNA from the plasmids was first quantified by a fluorometric method based using the Quanti-iT PicoGreen dsDNA assay kit (ThermoFisher, USA). Reactions for fragmentation, end repair, A tailing, adaptor ligation and post ligation clean up were set up as outlined below. The plasmids were first subjected to different fragmentation incubation times of 3, 5, 7 and 10 minutes to determine an optimal incubation time that yielded DNA fragments in the range 550 to 600 bp. Analysis of the fragmented DNA from the varied time experiments were done using a bioanalyser (Agilent 2100 analyser). The optimal fragmentation time was determined to be 9 minutes, which gave about 550 bp to 600 bp fragment length. After fragmentation, the fragments were enzymatically end repaired and A-tailed to prepare

them for adaptor ligation. The DNA fragments were ligated to paired end (PE) Illumina adaptor sequences, index1 (i7) adaptors and index2 (i5) adaptors. Post-ligation clean-up was also done using SPRI on Agencourt AMPure XP beads (Beckman Coulter, USA). Post-library preparation and another quantification of the libraries were done using the Quanti-iT PicoGreen dsDNA assay kit (ThermoFisher, USA). Each library was pooled into one tube based on the determined concentrations. For HPV211-HPV216, 72 clones were multiplexed together into one Illumina library and for HPV219-HPV222, 68 clones were multiplexed into a second run. In total 140 clones were sequenced (Table 2.3). Sequencing was done using Illumina MiSeq 300 bp paired end sequencing at Macrogen in Korea (Seoul, South Korea). Each library was sequenced on a separate run.

2.1.7 ILLUMINA DATA QUALITY CONTROL

A total of 32,409,876 reads, with a median of 578,992 reads per sample (range: 352,306-784,098) were obtained for each HPV plasmid. Illumina sequence reads were processed in CLC Genomic Workbench (Qiagen, Germany). Reads were trimmed using a quality score limit of 0.05 and any reads with more than 2 ambiguities discarded (Figure 2.2). *De novo* assembly of reads was done in CLC genomics.

2.1.8 GENOMIC CHARACTERISATION AND PHYLOGENETIC ANALYSIS

Sequence trimming, alignment and the *de novo* genome assembly was done in CLC Genomic Workbench (GW) v8.5.1 (Qiagen, USA). Genome lengths and nucleotide positions of the different ORFs were determined in CLC GW. The GC content was computed using an online tool (http://www.biologicscorp.com/tools/GCContent). Multiple sequence alignments (MSAs) were done in CLC GW and Multiple Sequence Comparison by Log-Expectation

(MUSCLE) (Edgar, 2004b). Variant, nucleotide and amino acid mismatches among clones of HPV212, HPV213, HPV215, HPV220 and HPV222 were analysed using nucleotide mismatches highlighter (Keele et al., 2008). Pairwise comparisons of variants of the same novel HPV type and for each putative novel HPV type with its most closely related known HPV, were also done in CLC GW.

ORF prediction was done in CLC GW (Qiagen, Germany). The procedure to annotate the novel HPV genomes is illustrated in Figure 2. 1. The initial step was the prediction of all possible ORFs in the forward strand of each complete novel genome in CLC GW. A minimum ORF nucleotide length was set at 300 nucleotides. The predicted ORFs were then compared to those of the closest known HPV types as annotated in the PaVE (Papillomavirus Episteme) database. The potential ORFs were then translated and subjected to BLAST in NCBI protein database to check if they relate to the closest relative of the novel type. All the novel types were checked to have the backbone ORFs E1, E2, L1 and L2 encoded by all PVs, some HPVs lack E6 or E7 and most Gamma-HPVs have been shown to lack E5. A similar but not identical manual annotation has been described elsewhere (Van Doorslaer et al., 2013b). The final step was a multiple sequence alignment of all the homologous novel HPV proteins with that of their closest known relatives to check for expected length and identification of known conserved regions (Figure 2.1). The first position of the complete genomes of HPV211, HPV212, HPV213, HPV215, HPV216, HPV219, HPV220, HPV221 and HPV222 were set at the first ATG of the E6 ORF. For HPV214, the first position was set at the first ATG of the E7 ORF because it lacks the E6 ORF. E1^E4 and E2^E8 splice donor/acceptor sites were predicted using manual inspection of splice sites predicted by NNSplice 0.9 (Reese et al., 1997), filtered using the set of criteria defined by Van Doorslaer and co-workers (Van Doorslaer et

al., 2017a). These criteria were the result of rigoroud comparative genomics among well known PVs. For example, the E4 portion of the E1^E4 mRNA is derived from the +1 reading frame of E2. Similarly, the E8 fragment of the E8^E2 protein is contained within the E1 +1 frame. E1^E4 and E8^E2 utilize the same splice acceptor site located within E2 (Van Doorslaer et al., 2017a). A checklist to ensure the spliced gene products were accurately predicted is outlined in Supplementary Table 2.9 in the appendix.

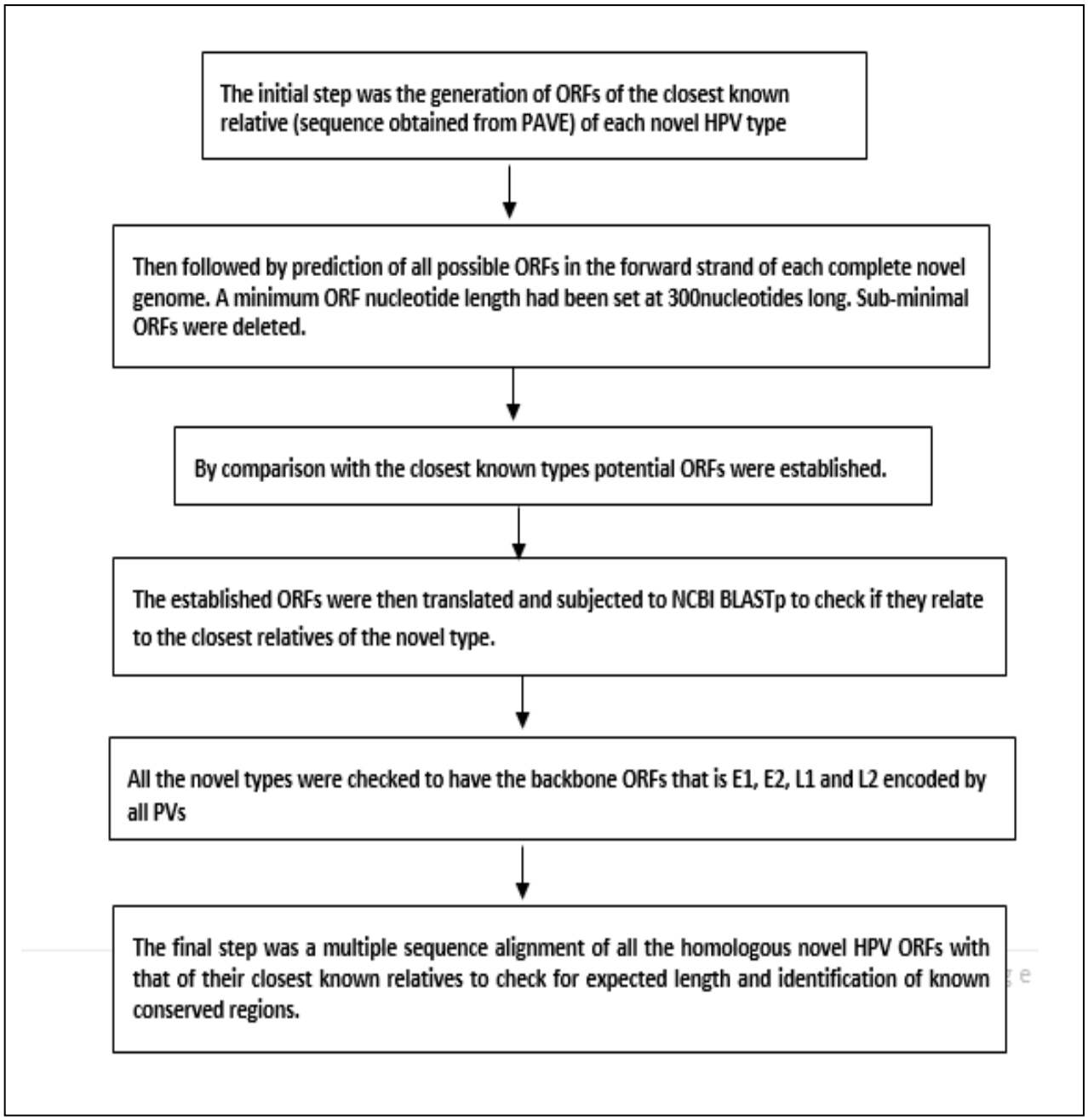


Figure 2.1 Genome annotation procedure.

Table 2.3 Summary of work done on the amplification, cloning, Sanger and Illumina sequencing of novel Gamma-HPV types.

Original Putative Novel Type	Closest HPV Match	Genome amplified	Genome cloned	Number of clones generated	Sanger sequence confirmation	Illumina Miseq	*De novo* assembly	Final number of clones analysed	Assigned HPV Reference Centre number
CT01	HPV153	✓	✓	11	✓	✓	✓	10	HPV219
CT02	HPV168	✓	✓	11	✓	✓	✓	11	HPV211
CT03	HPV144	✓	✓	11	✓	✓	✓	11	HPV212
CT04	HPV153	✓	✓	10	✓	✓	✓	10	HPV213
CT05	HPV146	✓	✓	12	✓	✓	✓	10	#
CT06	HPV103	✓	✓	8	✓	✓	✓	8	HPV214
CT07	HPV129	✓	✓	8	✓	✓	✓	8	HPV215
CT08	HPV144	✓	✓	10	✓	✓	✓	10	HPV220
CT09	HPV142	✓	✓	10	✓	✓	✓	10	HPV221
CT10	HPV144	✓	✓	9	✓	✓	✓	0	X
CT11	HPV154	✓	✓	9	✓	X	X	0	X
CT12	HPV129	✓	✓	11	✓	✓	✓	11	HPV216
CT13	HPV175	X	X	0	X	X	X	0	X
CT14	HPV101	X	X	0	X	X	X	0	X
CT15	HPV95	✓	✓	10	✓	✓	✓	5	HPV222
Fa101/Fa55/Fa91/SE55		✓	✓	20	✓	✓	X	0	X
				150		**140**		**104**	

X-Procedure was not successfully done or was not done due to an unsuccessful prior procedure.

✓-Procedure successfully done.

#-HPV type number had been previously assigned i.e. CT05 was not a novel HPV type

Comments: In total 17 full genomes were successfully amplified using Long range PCR and these were also successfully cloned. A total of 150 clones were generated using p-GEM T easy and TOPO XL cloning systems, and 140 (72 in the first run and 68 in the second run) were selected for Illumina sequencing, 104 clones were finally analysed after trimming assembly and all quality checks.

2.1.9 CLONE SELECTION

In order to select the most suitable clone for submission to the International HPV Reference Centre for purposes of nomenclature, multiple whole genome sequence alignments of the clones of each novel type were done in CLC genomics, where the reference sequence was the consensus of the clones. Mismatches between the different clones were identified using an online single nucleotide polymorphism (SNP) highlighter from Los Almos data base (Keele et al., 2008). Relatedness of the clones of the same novel HPV type was determined from the figures generated by comparing the positions of the mismatches along the whole genome. The clone with the highest percentage pairwise identity to the consensus sequence was treated as the prototype reference sequence. Chosen prototypes were carefully analysed and the ORFs annotated, translated into amino acids sequences and pairwise compared to the other clones of the closest known HPV types to verify they were the right clones. The prototype reference clones were sent to the International HPV Reference Centre, which assigned HPV numbers on verification of the sequence (HPV211 to HPV216 and HPV219 to HPV222).

2.1.10 NUCLEOTIDE ACCESSION NUMBERS

The novel HPV DNA sequences of the prototypes were deposited in Genbank under the following accession numbers: HPV211 MF509816, HPV212 MF509817, HPV213 MF509818, HPV214 MF509819, HPV215 MF509820, and HPV216 MF509821. The GenBank accession numbers for HPV219, HPV220, HPV221, HPV222 genomes are MH172376, MH172377, MH172378, MH172379, respectively.

2.1.11 PHYLOGENETIC ANALYSIS

For the phylogenetic analysis, L1 nucleotide sequences from HPV211-to HPV216 and HPV219 to HPV222 were aligned with L1 sequences from the PaVE database (Van Doorslaer et al., 2013a) using MUltiple Sequence Comparison by Log- Expectation (MUSCLE) (Edgar, 2004a). A maximum likelihood tree was then generated with PhyML 3.0 (Guindon et al., 2010) using the GTR+I+R substitution model, as determined by jmodeltest (Darriba et al., 2012). The approximate likelihood ratio test (aLRT) was used to estimate branch support (Anisimova and Gascuel, 2006). The tree was visualised in Interactive tree of life (iTOL) 4.2.3 (Letunic and Bork, 2016), an online tool for the display and annotation of phylogenetic and other trees.

2.1.12 NUCLEOTIDE AND AMINO ACID VARIATION

Nucleotide and amino acid variations in novel HPV types that showed differences in the clones generated were determined from multiple sequence alignments of the clones of each novel type. The percentage nucleotide and amino acid substitutions was also determined for each ORF. The selection pressures acting on the coding sequences of the novel HPVs were estimated by calculating codon-specific non-synonymous (d_N) and synonymous (d_S) substitution rates using the Synonymous Non-synonymous Analysis Program (SNAP) version 2.1.1 (Rodrigo, 2000) (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html). This program uses the unweighted pathway method of Nei-Gojobori (Nei and Gojobori, 1986) and the Jukes-Cantor model (Rodrigo, 2000). The d_N/d_S ratios were calculated, with a ratio of one indicating neutral selection, >1 diversifying positive selection and <1 negative or purifying selection (Chen et al., 2009).

2.1.13 IDENTIFICATION OF CONSERVED DOMAINS

The amino acid sequences of each novel HPV protein were aligned to homologous amino acid and nucleotide sequences of the closest known HPV types. These multiple sequence alignments with homologous proteins were done in CLC main workbench. Manual inspection of the MSA was used to identify putative conserved domains in the novel HPV types. Manual inspection in combination with the knowledge of the position of conserved domains in homologous proteins of known HPV types was used to verify the putative conserved domains. The protein sequences were manually inspected for the motifs identified in the literature as shown in Table 1.1.

2. 2 RESULTS

2.2.1 GENOMIC ORGANISATION

In this study ten novel HPVs were cloned into pGEM-T Easy vector/ TOPO XL vector and sequenced using Illumina MiSeq 300bp paired end sequencing. Eleven clones of HPV211, and another eleven of HPV212, ten clones of HPV213, eight clones of HPV214, eight of HPV215 and eleven of HPV216 were included in the first sequencing run. In the second run HPV219, HPV220 and HPV221 had ten clones each while HPV222 had 5 clones that were pooled for deep sequenced. Notnot all of the clones had useful sequences post-NGS especially after quality checks and trimming. The assembled sequences of the viral genomes revealed sizes ranging from 7096 bp for HPV213 to 7357 bp for HPV214 (Table 2.4). The ORF positions on the genome and sizes of the corresponding proteins are given in Table 2.4. The genomic organisation of HPV211, HPV213, HPV215, HPV216, HPV219 and HPV220 were typical of Gamma-HPVs, encoding five early (E1, E2, E4, E6 and E7) and two late (L1 and L2) proteins, and lacking the E5 gene. HPV212, HPV221 and HPV222 did not have an E4 start

codon as has been reported elsewhere (Li et al., 2012). The spliced E1^E4 transcript, that encodes the primary E4-gene product (Doorbar, 2013) was however identified in all ten genomes. HPV214 additionally lacked the ORF encoding the E6 protein but a putative ORF called E10, upstream of the E7 ORF (Figure 2.2) was identified (Van Doorslaer and McBride, 2016, Van Doorslaer et al., 2017a). The genomic organisation for HPV214 and HPV220 are illustrated as examples in Figure 2.3. HPV220 resembles the typical Gamma-HPVs genome organization and HPV214 shows a lack of E6 ORF and the putative E10 ORF.

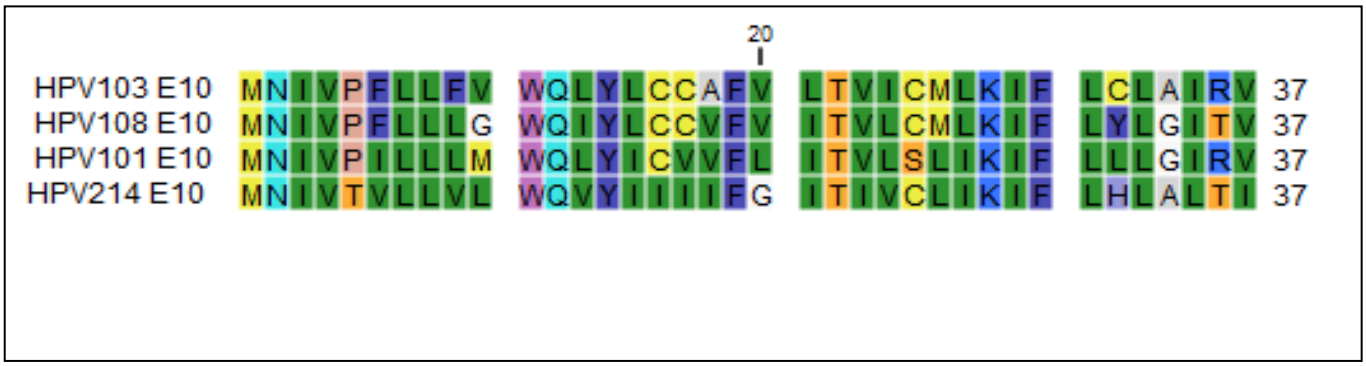


Figure 2.2 Alignment of putative E10, (37amino acid protein) of HPV214 with the E10 proteins of the most closely related HPV types from Gamma-6 species.

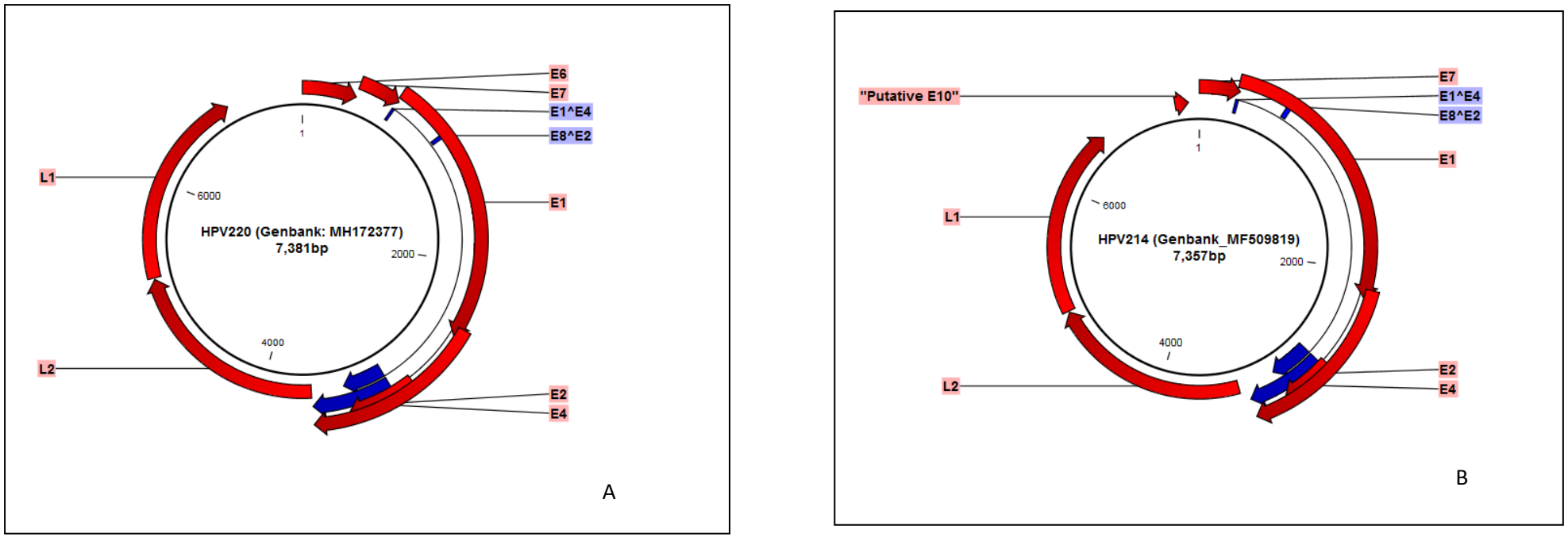


Figure 2.3 Examples of the genomic organization of the novel HPV types.

A) HPV220 resembling the typical Gamma-HPVs genome organization and B) HPV214 showing lack of E6 ORF and the putativeE10 ORF.

Table 2.4 Genome lengths of novel HPV types and ORF positions on the genome and sizes of proteins.

	HPV211	HPV212	HPV213	HPV214	HPV215	HPV216	HPV219	HPV220	HPV221	HPV222
Genome length (bp)	**7253**	**7208**	**7096**	**7357**	**7186**	**7233**	**7108**	**7381**	**7326**	**7275**
ORF position (Protein size in amino acids)										
E6	1-420 (140)	1-447 (149)	1-453 (151)	#	1-438 (146)	1-459 (153)	1-435(145)	1-423(141)	1-432 (144)	1-423 (141)
E7	417-707 (97)	444-743 (100)	422-715 (98)	1-303 (101)	440-733 (98)	462-755 (98)	422-715 (98)	420-719 (100)	434-727 (98)	420-710 (97)
E1	694-2508 (605)	727-2526 (600)	702-2513 (604)	284-2188 (635)	717-2528 (604)	739-2550 (604)	702-2513 (604)	703-2511 (603)	714-2546 (611)	700-2502 (601)
E2	2429-3619 (397)	2465-3628 (388)	2449-3624 (392)	2130-3296 (389)	2473-3624 (384)	2495-3667 (391)	2449-3621 (391)	2450-3616 (389)	2488-3672 (395)	2435-3631 (399)
E4	2940-3380 (147)	*<3036-3386 (117)	2921-3397 (159)	2704-3057 (118)	2900-3391 (164)	2922-3434 (171)	2921-3385 (154)	2919-3371 (151)	*<3053-3433 (127)	<*3030-3389 (120)
L2	3622-5220 (533)	3628-5154 (509)	3629-5122 (498)	3360-4973 (538)	3626-5137 (504)	3669-5186 (506)	3634-5127 (498)	3623-5224 (534)	3674-5245 (524)	3648-5180 (511)
L1	5231-6781 (517)	5166-6719 (518)	5133-6659 (509)	4984-6519 (512)	5146-6696 (517)	5195-6748 (518)	5138-6664 (509)	5237-6781 (515)	5256-6818 (521)	5192-6730 (513)

 # No E6 ORF ***start codon not determined**

2.2.2 CONSERVED DOMAINS ANALYSIS

All the ten novel HPVs had Zinc finger Binding Domains in the E6 and E7 proteins. In the E6, there were two Zinc finger Binding Domains separated by 36 amino acids, with ($CxxC(x)_{29}CxxC$) conserved sequence as described elsewhere (Latsuzbaia et al., 2018, Brancaccio et al., 2017, Dutta et al., 2017), except for HPV214 that did not have the E6 ORF. The E7 of all the novel types all had a single Zinc finger Binding Domain (Table 2.5).

Putative PDZ binding domains were identified in the E6 N-terminal of all the novel types except for HPV212, HPV220, HPV221, HPV222 and HPV214 (which lacks the E6 protein). The classical x(T/S)x(L/V) PDZ motif (Table 2.5), has also been described elsewhere (Martin et al., 2014, de Villiers and Gunst, 2009).

The LxCxE consensus motif (Ostrbenk et al., 2015, Mitsuishi et al., 2013, Wang et al., 2010) essential for binding of pRB was present only in the E7 protein of HPV214 and HPV222 (Table 2.4), this has also been described elsewhere (Dutta et al., 2017). The E1 protein of all the novel types contained the ATP binding sites with the typical consensus motif $G(x)_4GK(T/S)$ in the C-terminus (Chen et al., 2007b, Latsuzbaia et al., 2018). NLS sites were identified in the E1, E2, L1 and L2 proteins of all the novel HPV types. The NLS sites had varying consensus motif sequences depending on which protein they were found (Table 2.5), these classical NLS sites have been described by other authors (Lange et al., 2007). The nuclear export signals were identified in the E1 protein and have been previously described (Bergvall et al., 2013a). The DNA recognition motif was found in the E2 protein of all the novel HPV types, but HPV214 has a slightly different motif from the $GxxNx\mathbf{L}KCxRxR(x)_8$ (Graham, 2016), with an Leucine to Threonine modification and the functionality of this

motif can only be investigated by functional assays. We observed a typical transmembrane binding domain in the L2 of the novel HPV types $G(x)_3G(x)_3G$ (Wang and Roden, 2013, Bronnimann et al., 2013). Figure 2.4 shows domains in the E6 protein, a comprehensive outline of multiple sequence alignments of the ten novel HPV types' proteins and those of their closest known relatives is demonstrated in Supplementary Figures 2.16 to 2.21 in the appendix to chapter 2, also see Table 2.5 for a summary of the conserved domains.

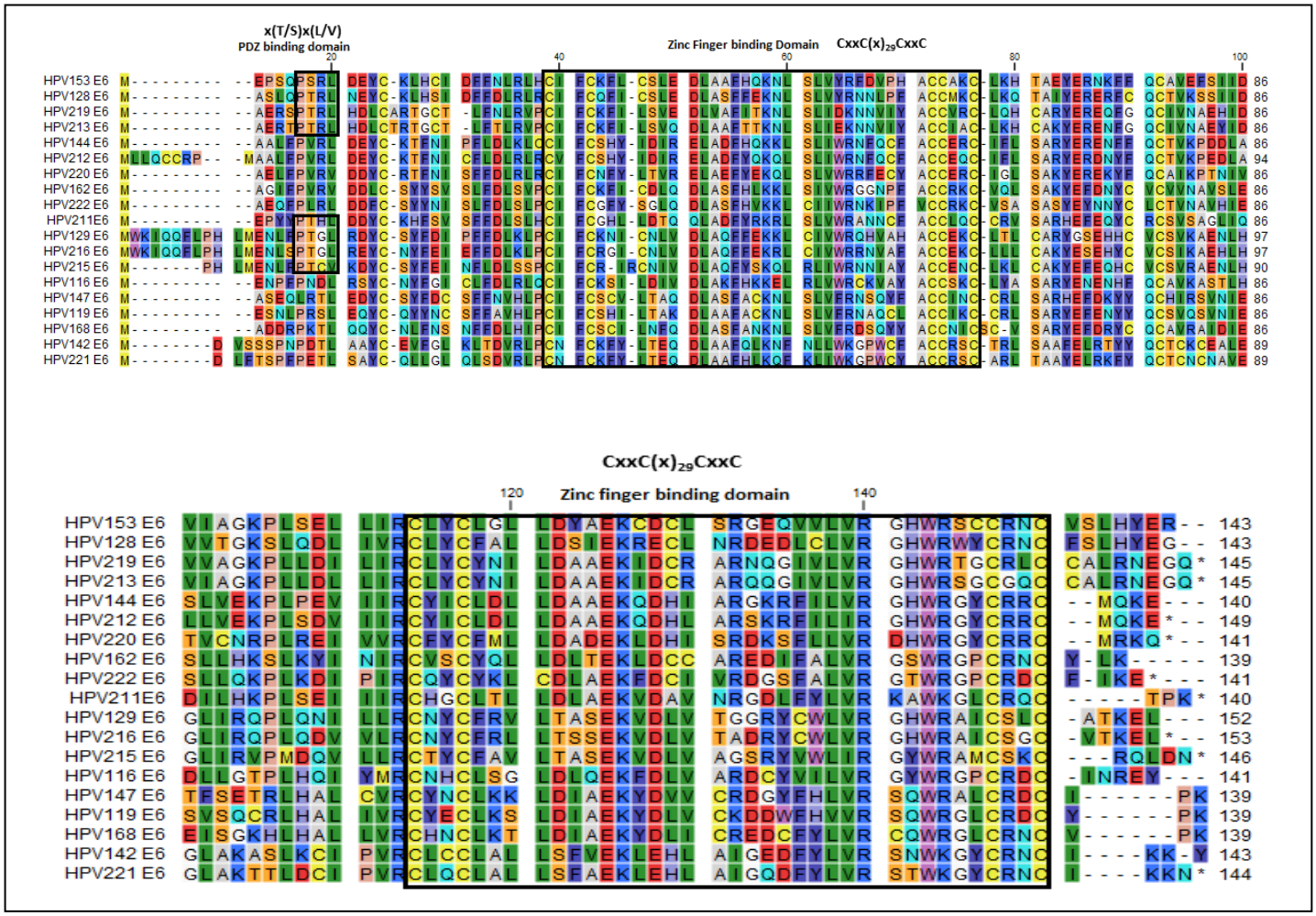


Figure 2.4 Alignment of the E6 proteins of the novel HPVs and closely related HPVs.

The positions of the PDZ binding domain and Zinc finger binding domains in E6 are indicated by the black boxes.

Table 2.5 Presence and frequency of potential conserved domains in the novel HPV types.

Genes/ region	Domain	Motif	HPV211	HPV212	HPV213	HPV214	HPV215	HPV216	HPV219	HPV220	HPV221	HPV222
E6	GATA Type Zinc fingers	$CxxC(x)_{29}CxxC$	2	2	2	NA	2	2	2	2	2	2
	PDZ binding domain	x(T/S)x(L/V)	1	-	1	NA	1	1	1	-	-	-
E7	E7 Conserved region 1		1	1	1	1	1	1	1	1	1	1
	GATA Type Zinc fingers	$CxxC(x)_{29}CxxC$	1	1	1	1	1	1	1	1	1	1
	pRB binding domain	LxCxE	-	-	-	1	-	-	-	-	-	1
E1	ATP binding site	$G(x)_4GK(T/S)$	1	1	1	1	1	1	1	1	1	-
	Bipartite nuclear localisation signal	KRK and KRRL	1	1	1	1	1	1	1	1	1	1
	Nuclear export domain putative	$(L/I)(x)_{2-3}(L/I)xx(L/I/V)x(L/I/V)$	1	-	1	1	1	1	1	1	1	1
E2	Leucine zipper domain	$Lx_6Lx_6Lx_6L$	-	-	-	-	-	-	-	-	-	-
	DNA recognition helix	$GxxNxLKCxRxR(x)_8$	1	1	1	1*	1	1	1	1	1	1
	Nuclear localisation domain	RKRxR/KRRR/KRXR	1	1	1	1	1	1	1	1	1	1
L1	Nuclear localisation like domain	$K(K/R)R(K/R)$	1	1	1	1	1	1	1	1	1	1
L2	Nuclear localisation like domain	$(K/R)_3R(K/R)$	1	1	1	1	1	1	1	1	1	1
	Transmembrane binding domain	$G(x)_3G(x)_3G$	1	1	1	1	1	1	1	1	1	1
	Furin cleavage site	$Rx(K/R)R$	1	1	1	1	1	1	1	1	1	1
	Early polyadenylation site	$AAT(A)_3$	1	-	1	1	1	1	1	1	1	1
LCR	E2 binding sites	$ACC(N)_6GGT$	3	3	1	3	3	2	1	2	2	2
	TATA binding box	$TAT(A)_3$	1	1	1	1	1	1	1	1	1	1
	Late polyadenylation site	$AAT(A)_3$	1	1	1	1	1	1	1	1	1	1

N represents any nucleotide, and x represents any amino acid.

Long control region (LCR),

Retinoblastoma binding protein (pRB)

*The DNA recognition motif of HPV214 E2 has a L to T modification $GxxNx\mathbf{T}KCxRxR(x)_{8)}$

2.2.3 SEQUENCE SIMILARITY AND PHYLOGENY

Pairwise comparison of all the novel types showed more than 60% identity to each other meaning they are members of the same genus but different species with the exception of HPV215 and HPV216 (72%) which belong to the same species, Gamma-9 (Bernard et al., 2010). In terms of the most closely related known HPV types, HPV211 was most closely related to HPV168 with a pairwise identity of 72% based on the L1 nucleotide sequence, HPV212 to HPV144 with a pairwise identity of 82.9%, HPV213 to HPV153 with a pairwise identity of 71.8%, HPV214 to HPV103 with a pairwise identity of 75.3%, HPV215 to HPV129 with a pairwise identity of 76.8% and HPV216 was to HPV129 with a pairwise identity of 79.2% (Table 2.6 and Figure 2.5). HPV211 clustered in the Gamma-8, HPV212 in the Gamma-17, HPV213 in the Gamma-13 and HPV214 in the Gamma-6 species, HPV215 and HPV216 clustered with HPV129 of the Gamma-9 species as shown in the maximum likelihood tree (Figure 2.5). HPV219 is phylogenetically most closely related to HPV213 (87% identity in L1 gene) of the Gamma-13 species, HPV220 to HPV212 (72%) of Gamma-17, HPV221 to HPV142 (80%) of Gamma-10, HPV222 to HPV162 (73%) of Gamma-19 (Table 2.6), as also shown in the maximum likelihood tree (Figure 2.5). The HPV genotypes share <90% identity in the L1 gene (Bernard et al., 2010), all ten viruses are therefore novel genotypes.

For all the putative ORFs, nucleotide and amino acid percentage pairwise identities to the closest recognised HPV types are shown in Table 2.6, and all were closely related to members of the Gamma-PVs genus.

Table 2.6 Nucleotide (amino acid) percentage pairwise identity to closest HPV type.

	HPV211	**HPV212**	**HPV213**	**HPV214**	**HPV215**	**HPV216**	**HPV219**	**HPV220**	**HPV221**	**HPV222**
HPV type most closely related (% identity)	HPV168 (72)	HPV220 (72)	HPV219 (87)	HPV103 (75.3)	HPV 129 (76.8)	HPV129 (79.2)	HPV213 (87)	HPV212 (72)	HPV142 (80)	HPV162 (73)
Gamma-species	8	17	13	6	9	9	13	17	10	19
E6	66 (51)	87 (89)	72 (56)	-	73 (69)	84 (83)	71(52)	73(62)	79(78)	72(66)
E7	70 (55)	88 (87)	73 (61)	74 (69)	77 (74)	80 (76)	76(61)	80(44)	91(88)	73(68)
E1	69 (61)	88 (92)	75 (71)	72 (65)	79 (77)	85 (88)	76(72)	73(69)	86(86)	75(70)
E2	67 (55)	90 (90)	72 (56)	70 (62)	75 (70)	82 (76)	71(55)	74(67)	85(82)	71(60)
E4	69 (44)	89 (81)	79 (40)	72 (63)	74 (60)	83 (75)	75(37)	75(62)	84(78)	74(60)
L2	65 (53)	80 (85)	70 (52)	71 (62)	69 (65)	77 (75)	68(46)	68(61)	80(85)	68(64)
L1	72 (71)	83 (92)	72 (68)	75 (75)	77 (80)	79 (87)	76(67)	69(77)	83(89)	71(71)

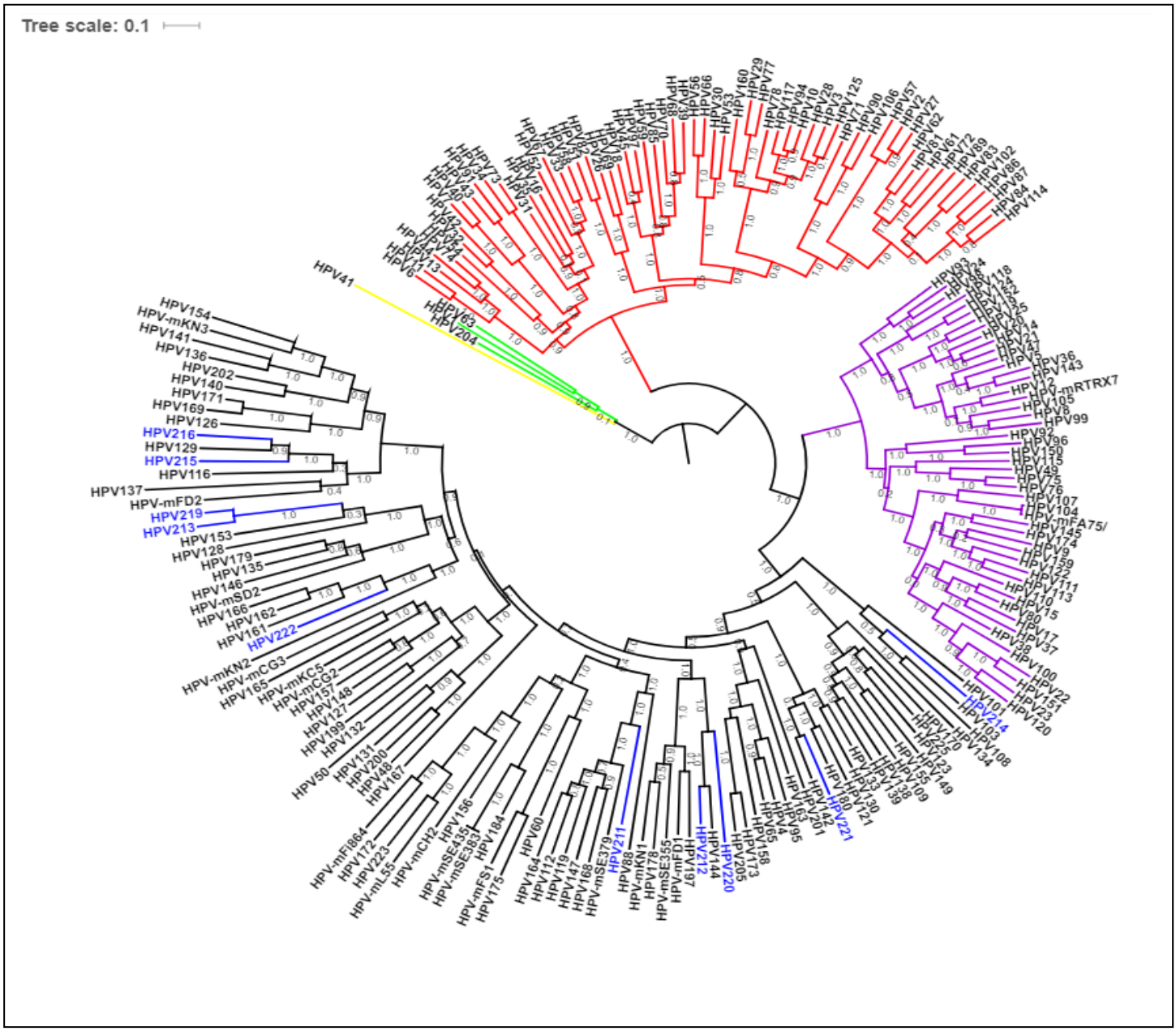


Figure 2.5 Maximum likelihood tree of the novel HPV types and related types.

Generated from Muscle alignment (Edgar, 2004a) of L1 nucleotide sequences in PHYML (Guindon et al., 2010) with GTR substitution. Approximate likelihood ratio test branch support values are indicated on the branch lines. Blue indicates the novel HPVs, black, yellow, red, purple and green branches represent the Gamma-HPVs, Nu-HPVs, Alpha-HPVs, Beta-HPVs and Mu-HPVs, respectively.

2.2.4 INTRA-SAMPLE VARIATIONS

All ten clones of HPV211, eight of HPV214, three of HPV216, ten of HPV219 and ten of HPV221 were identical and showed 100% pairwise identity. The ten clones of HPV213, and six of HPV215, and nine of HPV212, four of HPV220 and five clones of HPV222 were not 100% identical (but at least >99.5% identical). It was noteworthy that not all the clones that were sequenced had useful sequences especially after quality checks and trimming, hence the reduced number of clones per novel type during the analysis stage. Analysis of mismatches between the nine genomic clones of HPV212 showed a total of 67 positions that varied along the 7208 bp genome (Figure 2.6). All the HPV212 clones were unique. Analysis of mismatches between the 10 genomic clones of HPV213 showed a total of 51 mismatch positions that varied along the 7096 bp genome (Figure 2.5). HPV213 had 5 unique clones. The six genomic clones of HPV215 showed a total of 50 mismatch positions along a 7186 bp genome (Figure 2.6). HPV215 had 3 identical and 3 different clones (Figure 2.5). HPV220 had 4 different genomic clones that showed 17 mismatch positions along a 7381 bp genome. The 5 different clones of HPV222 showed a total of 24 mismatch positions along the 7275 bp genome (Figure 2.7).

Nucleotide changes that resulted in amino acid changes in the different ORFs are shown in Figure 2.8 and Figure 2.9. There were no non-synonymous changes in conserved functional domains except for two HPV213 clones that had a change from Cysteine to Phenylalanine at the last position of the E7 Zinc finger Binding Domain. In summary, percentage amino acid substitutions for HPV212 clones ranged from 0.98% in the L2 ORF to 2.67% in the E1 protein, for HPV213 clones from 0% in the E7 protein to 1.18% in the L1; HPV215 from 0.56% in the E4 to 1.79% in the L2, for HPV220 clones from 0.18% in L2 to 0.78% in L1 and lastly for

HPV222 clones amino acid substitution ranged from 0.5% in E1 to 0.78% in the L1 protein. Table 2.7 and Table 2.8 show that all the proteins of the novel HPV types had d_N/d_S ratios less than 1, which is indicative of negative selection pressure or purifying selection.

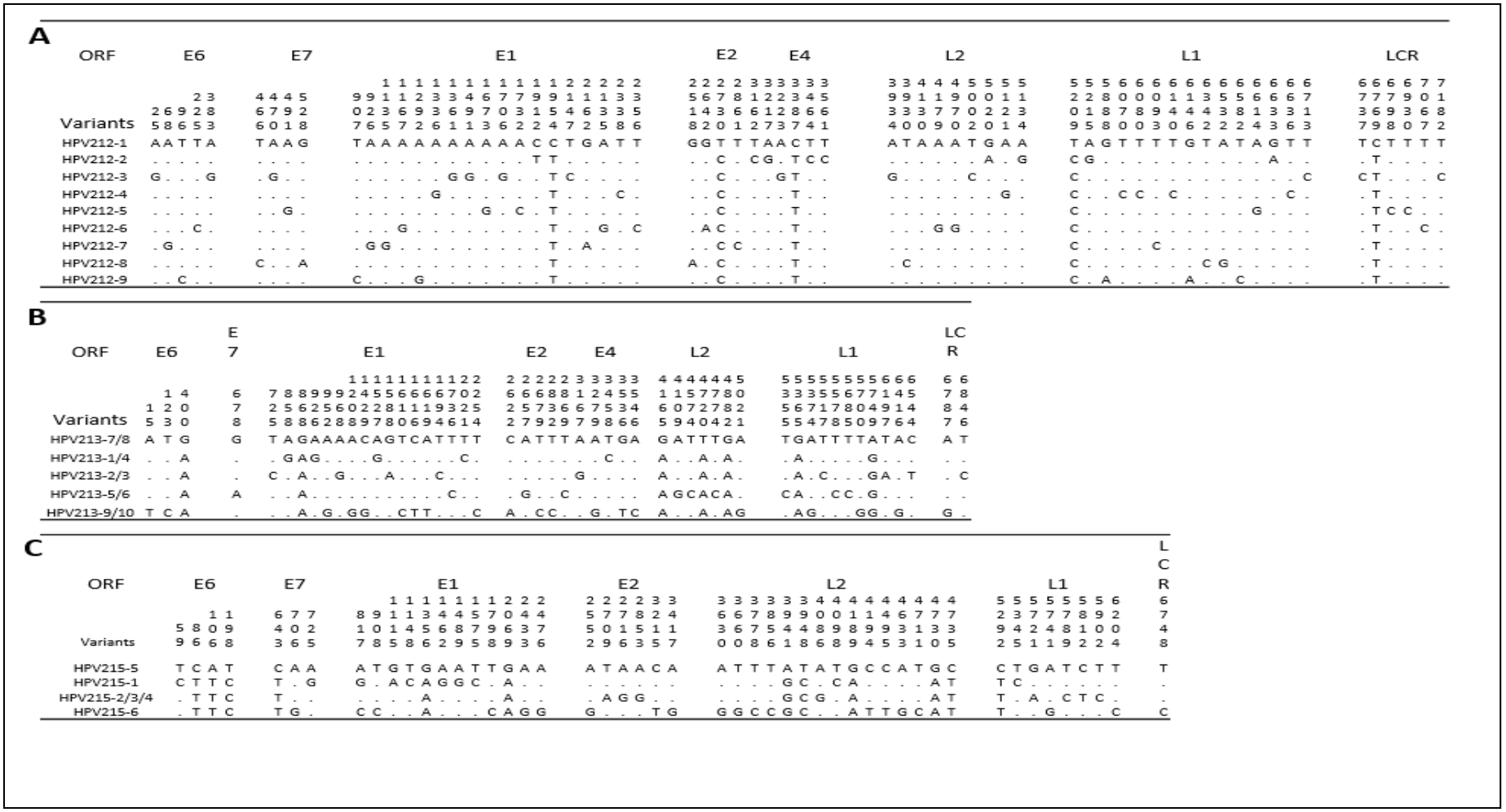


Figure 2.6 HPV212, HPV213 and HPV215 nucleotide mismatch positions.

Positions of mismatches are based on the prototype Reference sequences (HPV212-1, HPV213-7 and HPV215-5) deposited in Genbank.

	E7	E1			E2			L2		L1					LCR		
Variants	**470**	**1159**	**1577**	**2385**	**2561**	**2819**	**3202**	**3854**	**4763**	**5477**	**5673**	**5951**	**6290**	**6596**	**6868**	**6898**	**7042**
HPV220-5	C	A	T	A	A	G	A	T	T	T	A	A	T	C	C	T	C
HPV220-1	T	T	.	G	.	.	.	.	C	C	.	.	C	.	T	.	T
HPV220-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HPV220-3	.	.	C	.	G	A	G	C	.	.	G	G	.	G	.	C	.

	E1							E2				L2					L1							LCR
Variants	**947**	**1049**	**1287**	**1666**	**1965**	**2014**	**2242**	**2708**	**3108**	**3273**	**3531**	**3646**	**4333**	**4729**	**4738**	**4764**	**5482**	**5547**	**5593**	**6298**	**6442**	**6537**	**6552**	**6738**
HPV222-1A	A	A	T	G	.	A	G	A	T	A	A	A	A	T	G	C	A	A	C	G	T	.	T	A
HPV222-2C	.	.	.	.	.	G	.	.	.	G	G	.	.	A	A	T	G	G	T	A	.	.	.	.
HPV222-3G	G	G	.	.	A	.	.	.	.	G	.	G	.	.	.	T	.	.	.	A	C	.	C	.
HPV222-3E	.	.	.C	C	.	.	A	G	C	G	.	.	G	.	.	T	.	.	.	A	.	A	.	G
HPV222-2A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 2.7 HPV220 and HPV222 nucleotide mismatch positions.

Positions of mismatches are based on the prototype Reference sequences (HPV220-5 and HPV222-1) deposited in Genbank.

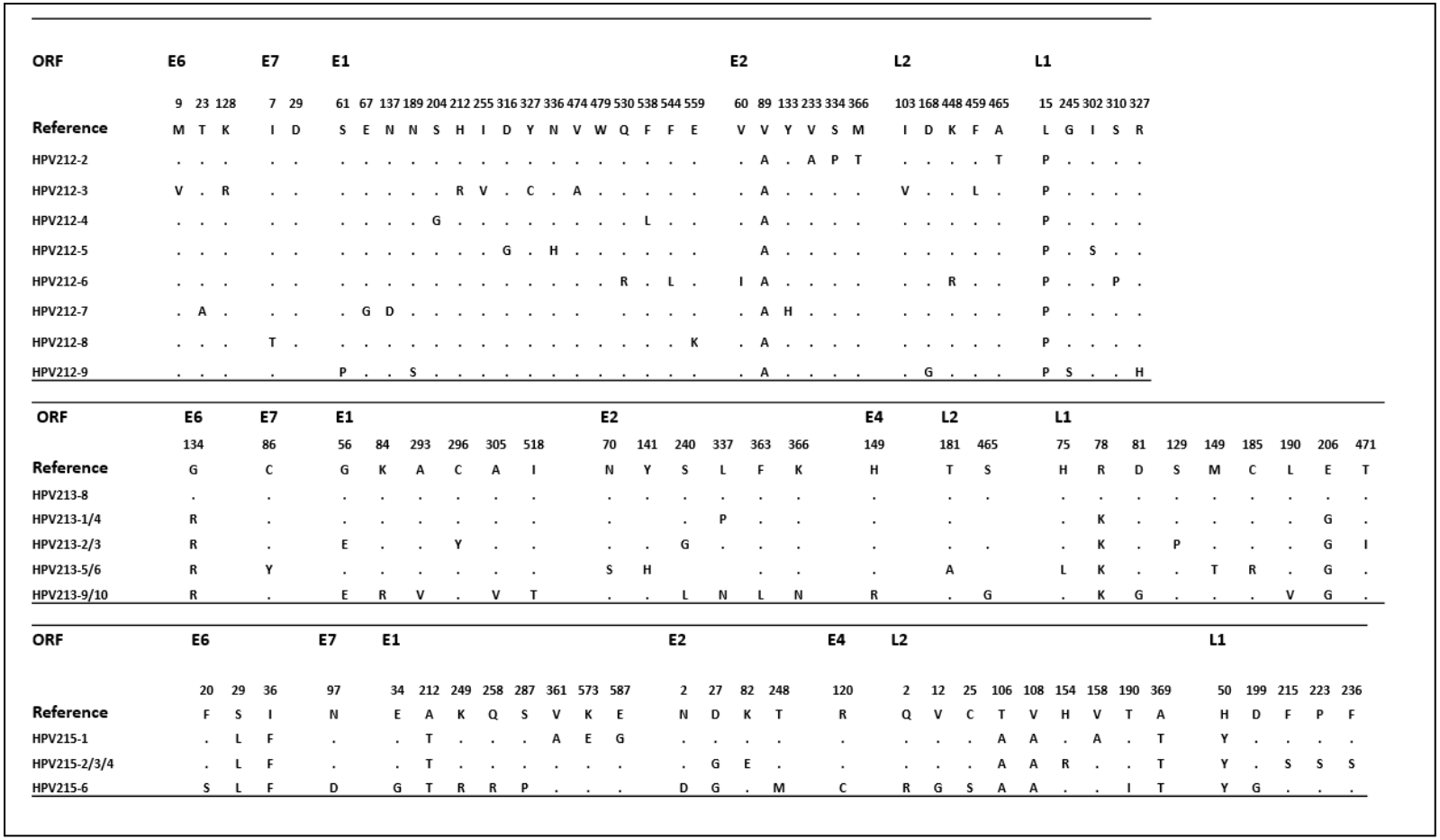
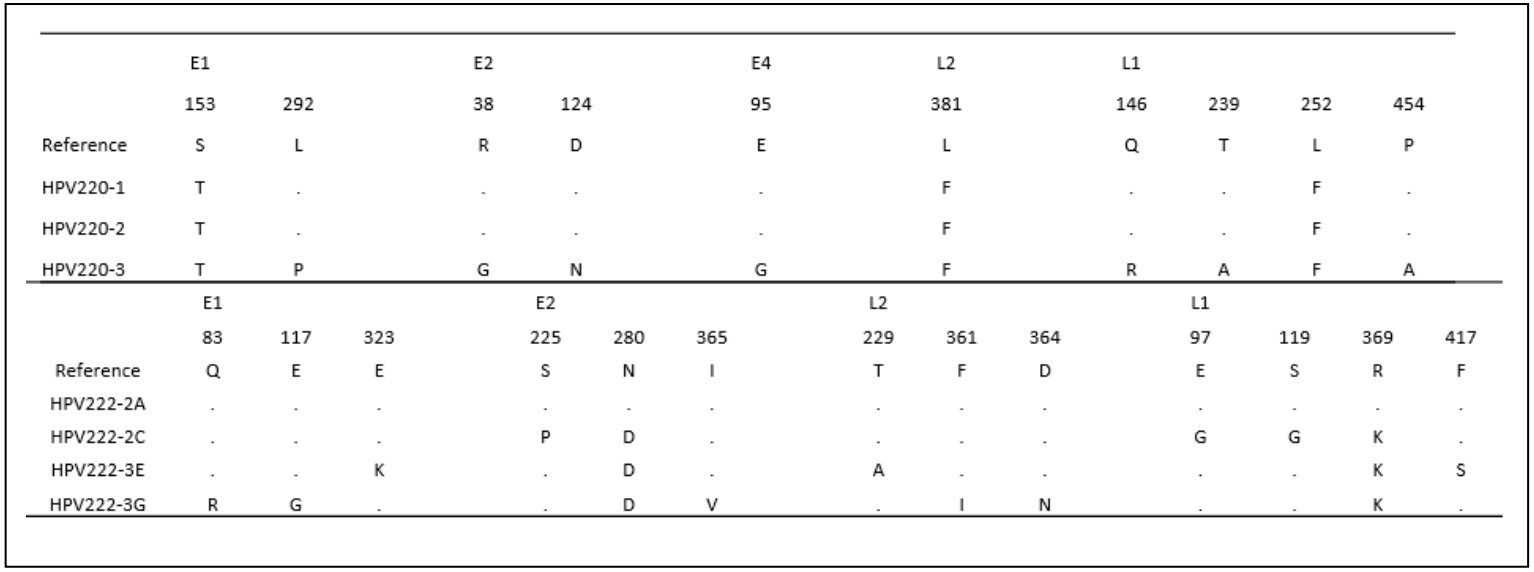


Figure 2.8 Amino acid variations in the predicted proteins of the HPV212 clones (with clone 1 as reference), HPV213 clones (with identical clones 7 and 8 as reference clones) and HPV215 clones (with clone 5 as the reference).



	E1			E2			E4	L2			L1			
	153	292		38	124		95	381			146	239	252	454
Reference	S	L		R	D		E	L			Q	T	L	P
HPV220-1	T	.		.	.		.	F			.	.	F	.
HPV220-2	T	.		.	.		.	F			.	.	F	.
HPV220-3	T	P		G	N		G	F			R	A	F	A

	E1			E2			L2			L1			
	83	117	323	225	280	365	229	361	364	97	119	369	417
Reference	Q	E	E	S	N	I	T	F	D	E	S	R	F
HPV222-2A	.	.	.	.	.	.	.	.	.	.	.	.	.
HPV222-2C	.	.	.	P	D	.	.	.	.	G	G	K	.
HPV222-3E	.	.	K	.	D	.	A	.	.	.	.	K	S
HPV222-3G	R	G	.	.	D	V	.	I	N	.	.	K	.

Figure 2. 9 Amino acid variations in the predicted proteins of the HPV220 clones (with lone 1 as reference), HPV222 clones (with identical clones 7 and 8 as reference clones).

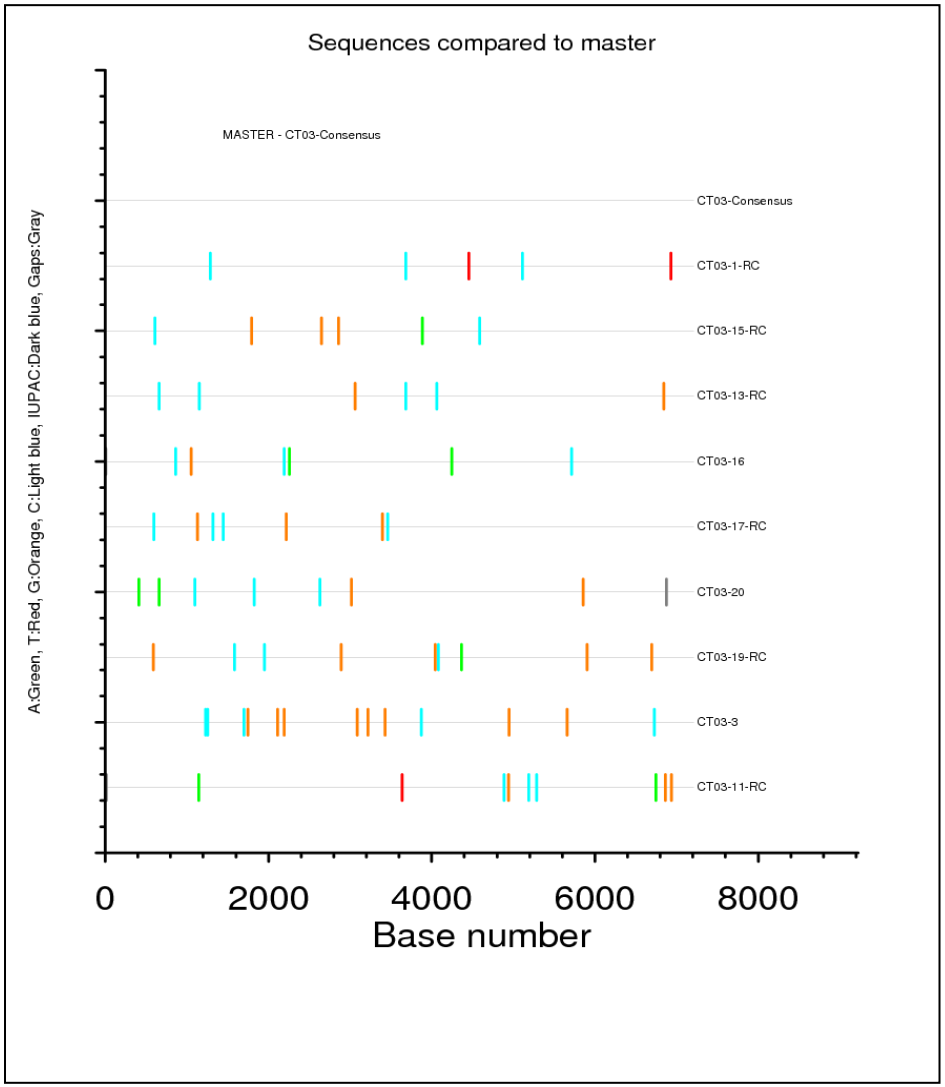
In total, 67 nucleotide mismatches positions along a 7208 bp were observed for HPV212, 51 mismatches positions for HPV213 along a 7096 bp and 50 mismatches for HPV215 along a 7186 bp length. Showing that all the HPV212 clones were different and hence the one closest to the consensus in identity was chosen as the "type" sequence. HPV212-1 was selected and annotated as shown in Figure 2.10. HPV213 clones are different (however, clone 9 and 10 were identical, as were 1 and 4; 5 and 6; 2 and 3 as well as 7 and 8). Hence the one closest to the consensus in identity was chosen. HPV213-7/8 were selected and annotated as the representative sequence for Genbank purposes as shown in Figure 2.10. HPV215 clones 2, 3 and 4 were similar and closer to the consensus, yet different from clones 1, 5 and 6. HPV215-5 was however selected as the representative sequence for Genbank purpose as also shown in Figure 2.11.

In total 17 nucleotide mismatches positions along a 7381 bp were observed for HPV220, 24 mismatches positions for HPV222 along a 7275 bp.

The nucleotide mismatch positions for HPV212, HPV213, HPV215, HPV220 and HPV222 were also analysed using an online single nucleotide highlighter (Keele et al., 2008) originally designed for identification of HIV mutations. The results of this analysis are shown in Figures 2.9 to 2.11. The mismatches do not pinpoint exact positions, but the Figures are intended to give a visual inspection of the amount of mismatched positions and how they are distributed along the sequences. These Figures used as visual representations, were also essential in visualizing the sequence closest to the consensus, which would then be chosen as the prototype reference sequence that was eventually send to the HPV International Reference Centre for nomenclature. HPV220 clones 5 and 2 were identical, yet different

from clones 1 and 3. HPV220-5/2 were however selected and annotated as the representative sequences for submission to Genbank as shown in Figure 2.11. HPV222 clones 1A and 2A were similar and closer, yet different from clones 3G, 2C and 3E. HPV222-1A was however selected and annotated as as the representative sequence for Genbank as shown in Figure 2.12.

Despite all the varying mismatch positions in the different sequences/variants of each clone, all the clones of each type showed between 99.5% to 100% pairwise identity to each other (Figure 2.13 and Figure 2.14). This is less than 0.5% difference in the sequences of the various clones indicating that they were not variants nor sub-lineages according to strict HPV nomenclature rules (Burk et al., 2013).



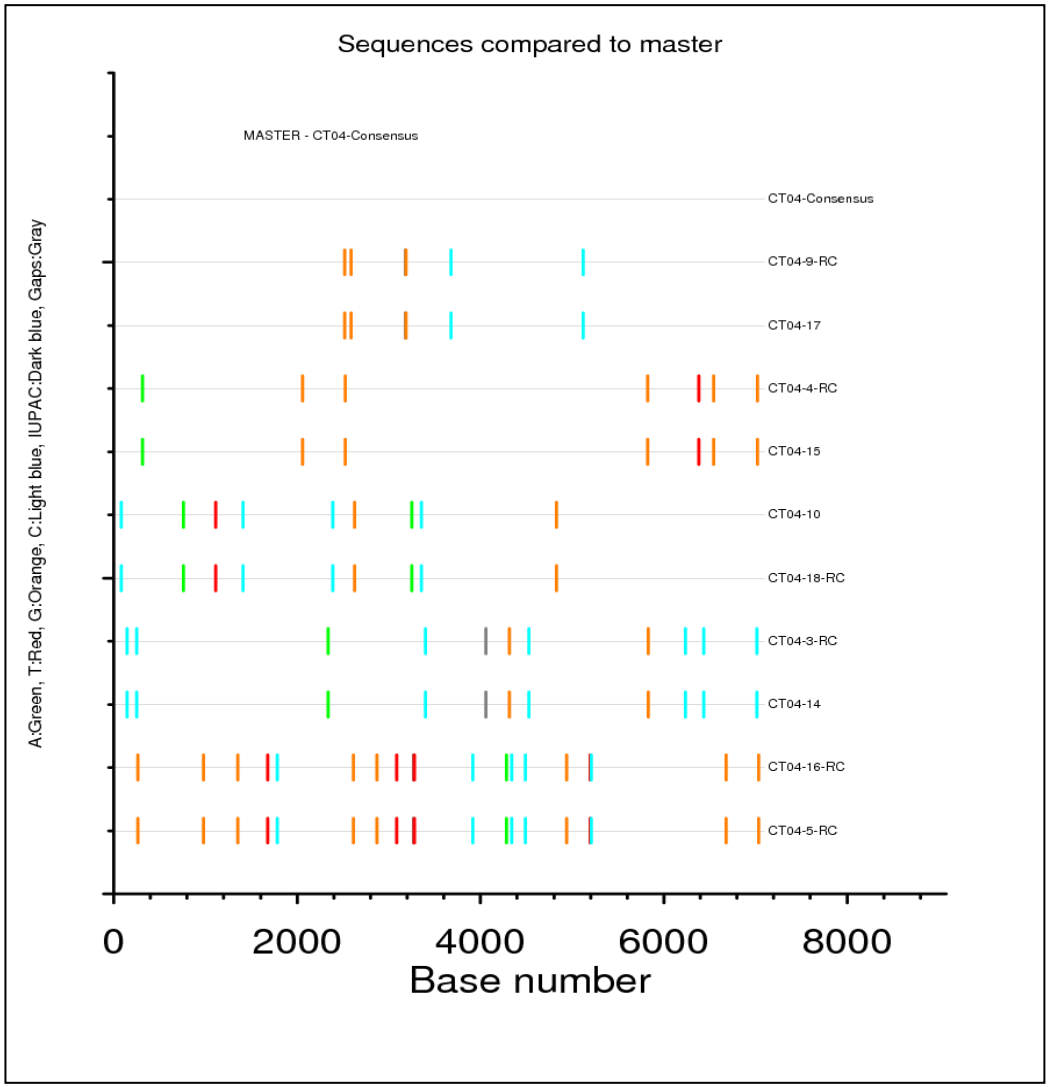


Figure 2.10 HPV212 (labelled as CT03 isolates) Clones aligned against the CT03 consensus sequence, HPV213 (labelled as CT04 isolates) Clones aligned against the CT04 consensus sequence.

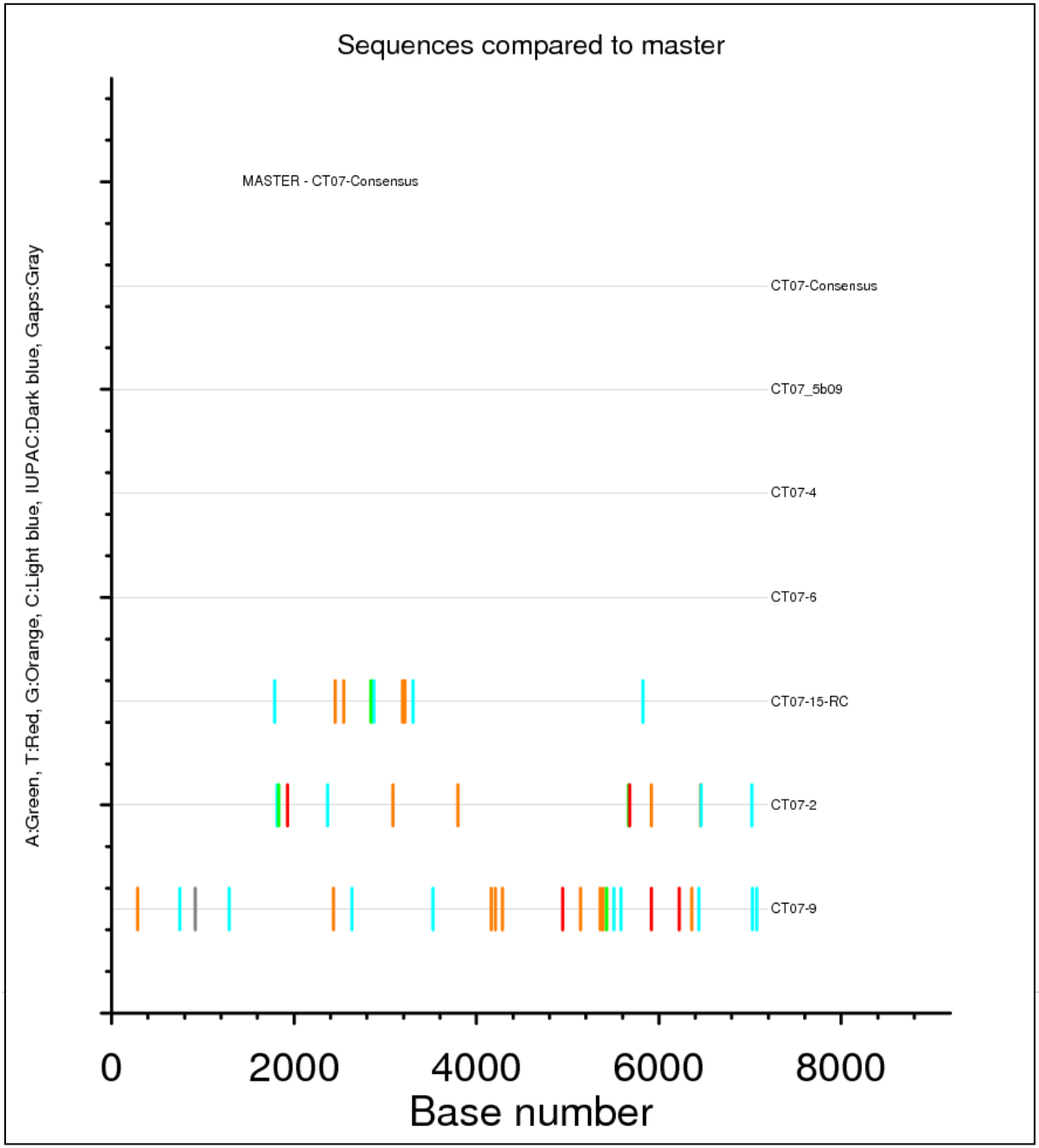


Figure 2.11 HPV215 (labelled as CT07 isolates) Clones aligned against the CT07 consensus sequence.

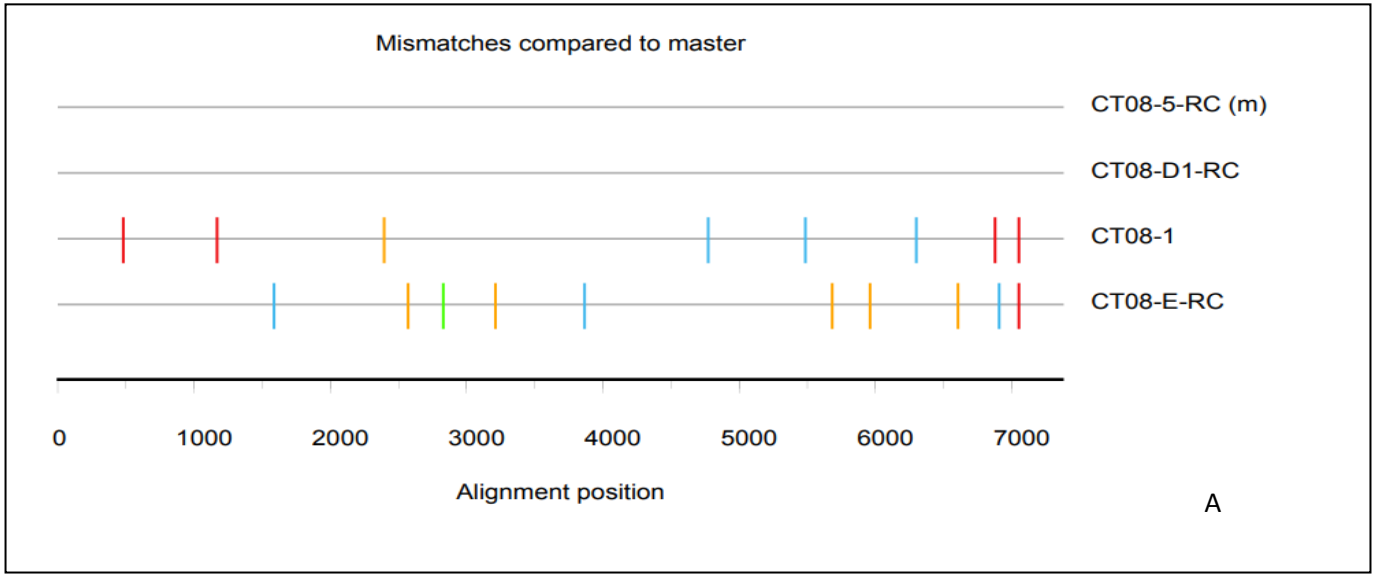


Figure 2.12 A) HPV220 Clones (labelled as CT08 isolates) aligned against the CT08-5 as master sequence.

B) HPV222 Clones (labelled as CT17 isolates) aligned against the CT17-1A as master sequence.

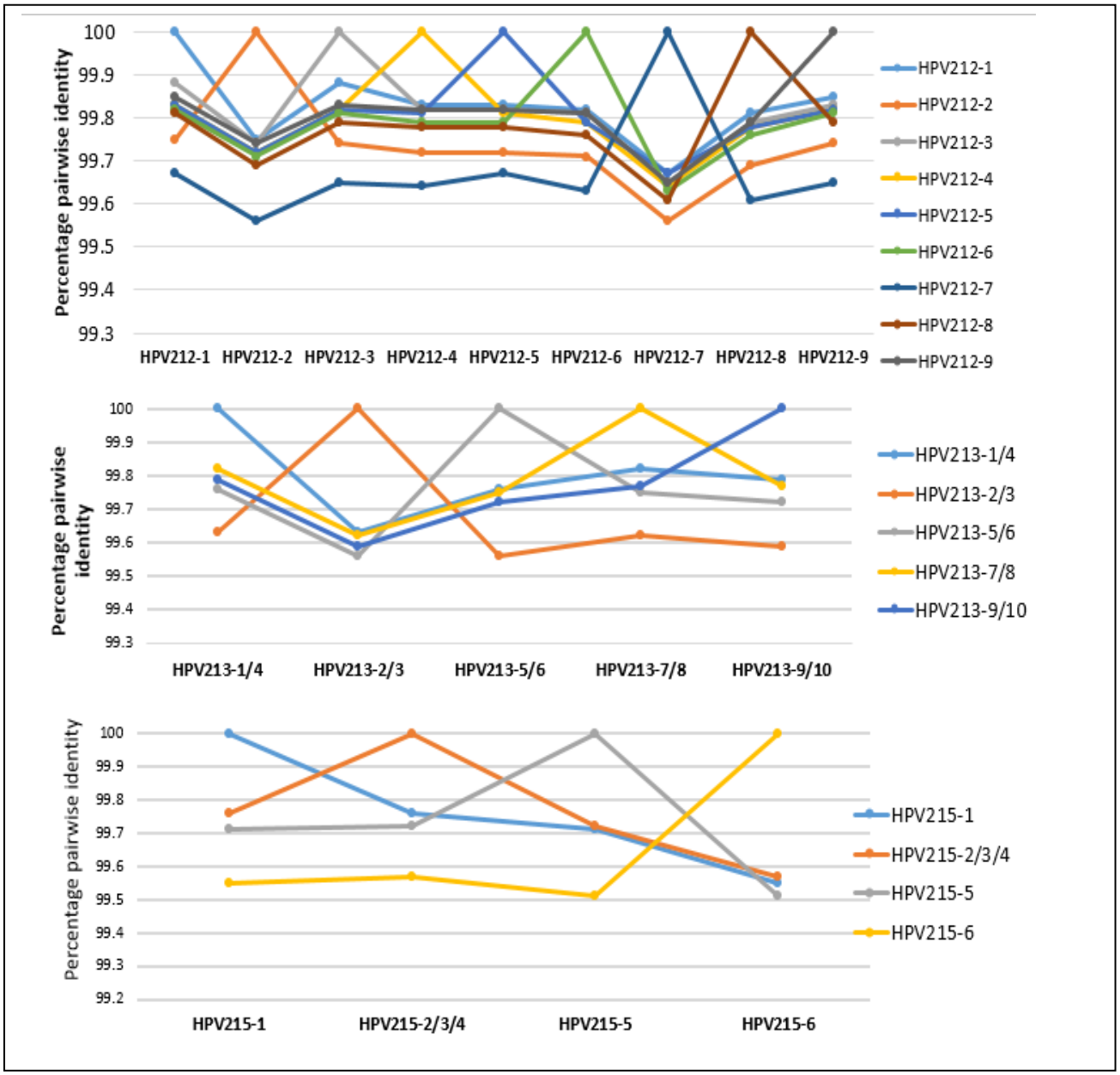


Figure 2.13 Percentage pairwise identities of the different clones of HPV212, HPV213 and HPV215.

Values of each pairwise identity of a given clone are connected by lines (colored differently for visual aid) and comparison to self is indicated by 100% pairwise identity point.

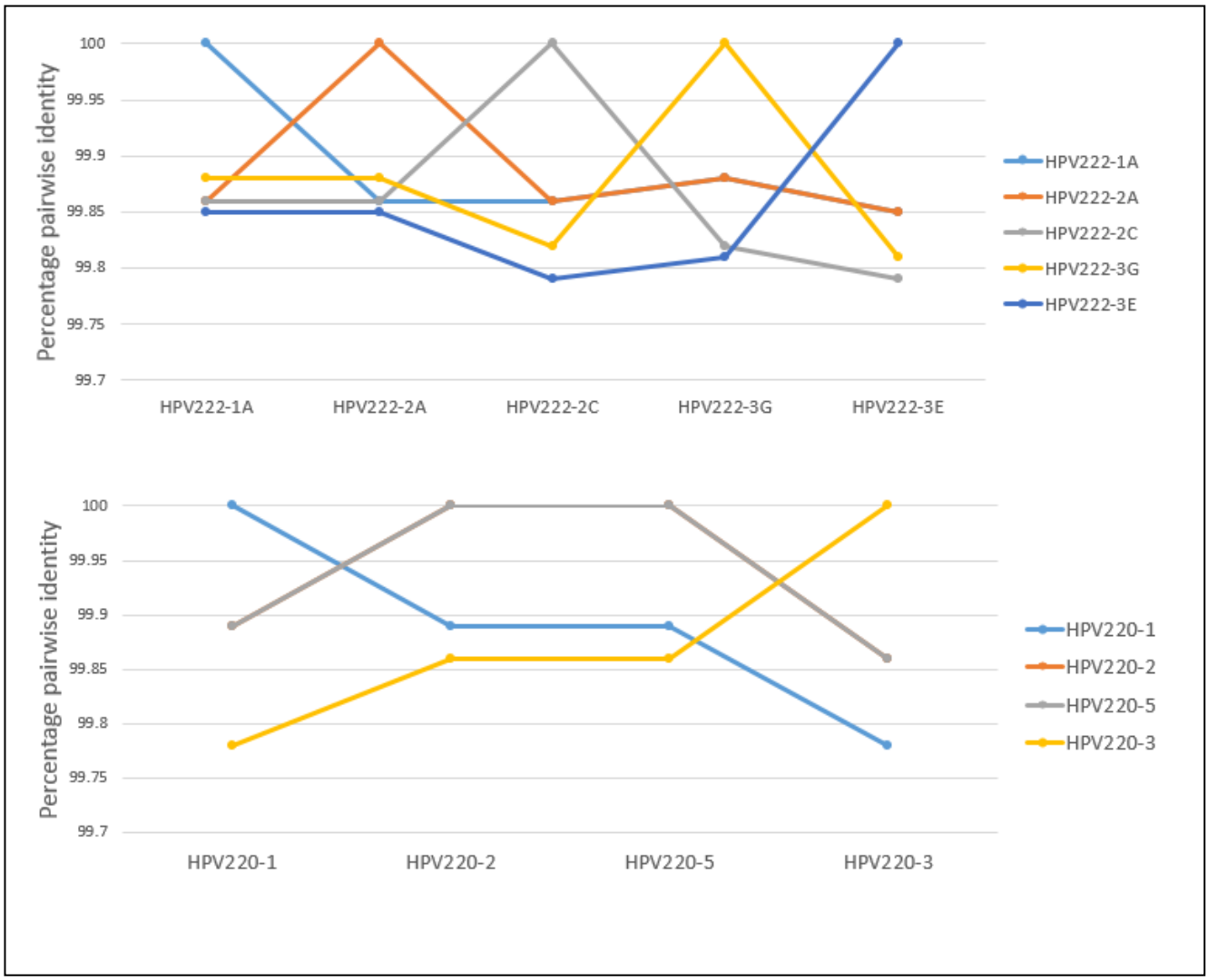


Figure 2.14 Percentage pairwise identities of the different clones of HPV220 and HPV222. Values of each pairwise identity of a given clone are connected by lines (colored differently for visual aid) and comparison to self is indicated by 100% pairwise identity point.

Table 2.7 Comparison of nucleotide and amino acid sequence variability and synonymous to non-synonymous changes within HPV212, HPV213 and HPV215 genes.

ORF	Number of Nucleotides	Number of variable Nucleotide positions	Number of amino acids	Number of nonsynonymous Changes (% aa substitution)	Number of synonymous changes(% aa substitution)	dN/dS ratio
HPV212						
E6	447	5 (1.12%)	149	3 (2.01%)	2 (1.34%)	0.47
E7	300	4 (1.33%)	100	2 (2.00%)	2 (2.00%)	undefined
E1	1800	19 (1.06%)	600	16 (2.67%)	3 (1.50%)	0.67
E2	1164	9 (0.77%)	388	6 (1.55%)	3 (0.77%)	0.27
E4	378	3 (0.79%)	126	3 (2.38%)	0 (0%)	undefined
L2	1527	9 (0.59%)	509	5 (0.98%)	4 (0.79%)	0.45
L1	1554	15 (0.97%)	518	6 (1.16%)	8 (1.54%)	0.45
URR	489	6 (1.23%)	-	-	-	
Total	7208	67 (0.92%)	2390	41 (1.72%)	22 (0.92%)	
HPV213						
E6	454	3 (0.66%)	151	1 (0.66%)	2 (1.32%)	0.14
E7	294	1 (0.34%)	98	1 (1.02%)	0(0%)	undefined
E1	1812	17 (0.94%)	604	9 (1.49%)	8(1.33%)	0.16
E2	1176	10 (0.85%)	392	6 (1.53%)	4 (1.02%)	0.2
E4	564	3 (0.53%)	188	1 (0.53%)	4 (2.13%)	0.54
L2	1494	7 (0.47%)	498	2 (0.40%)	5 (1.00%)	0.12
L1	1527	11 (0.72%)	509	9 (1.77%)	2 (0.39%)	0.55
URR	437	2 (0.46%)	-	-	-	
Total	7096	51 (0.72%)	2440	29 (1.19)	25(1.02)	
HPV215						
E6	438	4 (0.91%)	146	3 (2.05%)	1 (0.68%)	0.54
E7	294	3 (1.02%)	98	1 (1.02%)	2 (2.04%)	0.26
E1	1812	12 (0.66%)	604	8 (1.32%)	4 (0.66%)	0.49
E2	1512	6 (0.40%)	384	4 (1.04%)	2 (0.52%)	0.26
E4	534	1 (0.19%)	178	1 (0.56%)	0 (0%)	undefined
L2	1512	16 (1.06%)	504	9 (1.79%)	7 (1.39%)	0.39
L1	1551	8 (0.52%)	517	5 (0.97%)	3 (0.58%)	0.48
LCR	490	1 (0.20%)	-	-	-	

Table 2.8 Comparison of nucleotide and amino acid sequence variability and synonymous to non-synonymous changes within HPV220 and HPV222 genes.

ORF	Number of Nucleotides	Number of variable Nucleotide positions	Number of amino acids	Number of nonsynonymous Changes (% aa substitution)	Number of synonymous changes(% aa substitution)	dN/dS ratio
HPV220						
E6	423	0 (0%)	141	0(0%)	0(0%)	undefined
E7	300	1(0.3%)	100	0 (0%)	1(1%)	0
E1	1809	3(0.17%)	603	2(0.33%)	1(0.16%)	0.31
E2	1167	3(0.25%)	389	2(0.51%)	1(0.26)	0.59
E4	452	1(0.22%)	151	1(0.66%)	0 (0%)	undefined
L2	1602	2(0.12%)	534	1(0.18%)	1(0.18%)	0.33
L1	1545	5(0.32)	515	4(0.78%)	1(0.19%)	0.37
URR	600	3(0.5%)	-	-	-	
Total	7381	18(0.24%)	2433	10(0.41%)	8 (0.33%)	
HPV222						
E6	423	0 (0%)	141	0(0%)	0(0%)	undefined
E7	291	0 (0%)	97	0 (0%)	0 (0%)	undefined
E1	1803	7(0.39%)	601	3(0.50%)	4(0.67%)	0.59
E2	1197	4(0.33%)	399	3(0.75%)	1(0.25%)	0.34
E4	360	0 (0%)	120	0 (0%)	0 (0%)	undefined
L2	1533	5(0.33%)	511	3(0.59%)	2(0.39%)	0
L1	1539	7(0.46%)	513	4(0.78%)	3(0.58%)	0.24
URR	545	1(0.18%)	-	-	-	
Total	7275	24(0.33%)	2382	13(0.55%)	10(0.42%)	

Figure 2. 15 is a summary of the dN/dS ratios of all the ORFs of the five novel HPV types that showed variation among clones. A ratio of one indicates neutral evolution with values below one suggesting purifying selection (negative selection) and values more than one Darwinian selection (positive selection). It is apparent that all the ORFs show values below one. The E1, E2 and L1 across all the novel types showed some values at least above 0.2 although still below one.

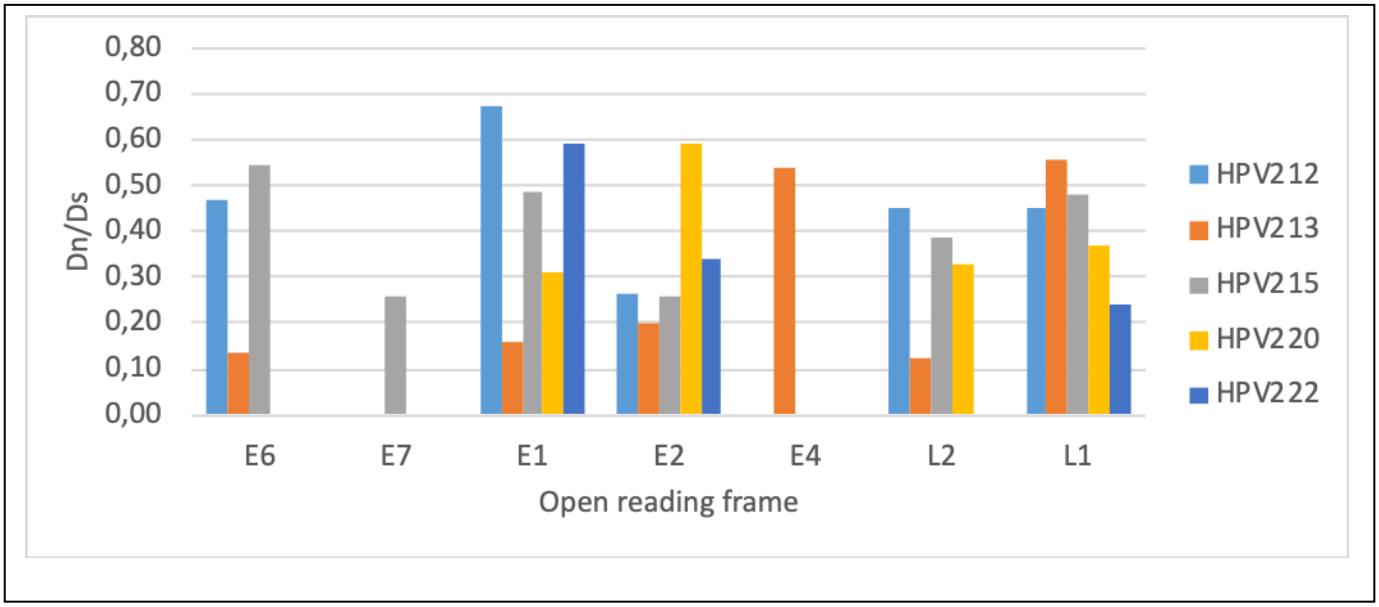


Figure 2.15 dN/dS ratios of each ORF of HPV212, HPV213, HPV215, HPV220 and HPV222.

2.3 DISCUSSION

This chapter describes the identification and characterisation of ten novel Gamma-HPVs. The amplification of whole genomes of these viruses from short L1 fragments using back to back primers in a long-range PCR using a touch down approach. Further described here is the cloning of the full genomes of the novel viruses and the application of NGS to construct by *de novo* assembly the genome sequences of these viruses. The application of several bioinformatics tools to characterise the viral sequences in terms of: genome organisation, conserved domains, phylogeny, genome sequence variation from other known HPVs and evolutionary selection pressures is also reported here. The curated sequences were then sent to the International Reference Centre in Karolinska, Sweden for verification and numbering. The ten novel viruses were confirmed as novel HPV types and assigned numbers HPV211 to HPV216 and HPV219 to HPV222 (http://www.nordicehealth.se/hpvcenter/reference_clones/). We deposited the sequences in Genbank under accession numbers: HPV211 MF509816, HPV212 MF509817, HPV213 MF509818, HPV214 MF509819, HPV215 MF509820, HPV216 MF509821, HPV219 MH172376, HPV220 MH172377, HPV221 MH172378 and HPV222 MH172379.

Most Gamma-PVs have been isolated from cutaneous lesions, but most of the novel types described in this thesis are related to Gamma-PVs that were also isolated from mucosal or anogenital tissue. HPV211 clustered with the Gamma-8 species alongside five other members. HPV212 became the second member of Gamma-17 species after the sole member HPV144 which was isolated from an oral rinse (Bottalico et al., 2011a). HPV213 became the third member of the Gamma-13 species after HPV153 which was isolated from a condyloma (Sturegard et al., 2013) and HPV128 which was isolated from a skin wart (Kohler et al., 2011) .

HPV215 and HPV216 became third and fourth members of the Gamma-9 species after HPV129 which was isolated from a skin wart and HPV116 which was isolated from a rectal swab (McLaughlin-Drubin and Munger, 2009b). HPV214 became the fourth member of Gamma-6 species after HPV101 and HPV103 which had been isolated from cervico-vaginal cells (Chen et al., 2007b) and HPV108 was isolated from a cervical lesion (Nobre et al., 2009b). HPV214 was also identified as HPVX by metagenomics sequencing of cervical DNA (Ameur et al., 2014b).

Future studies should examine whether Gamma-6 viruses can be identified in other mucosal and cutaneous sites. In addition to the lack of E6, it is apparent from the phylogenetic tree (Figure 2.4) that the four members of the Gamma-6 species, including the novel HPV214 described in this study, have a considerable phylogenetic distance from other Gamma-HPV types (see chapter 3 with predicted divergence times).

The E5 protein is only present in Alpha-HPVs and absent in Beta-, Gamma-, Mu- and Nu-papillomaviruses (Venuti et al., 2011). The functions of this 90 aa protein in high risk HPVs range from the binding of platelet derived growth factor (PDGF) and activating it, induction of cell transformation and MHC class 1 antigen presentation inhibition (thereby evading immune response) (Stanley et al., 2007). However, Gamma-HPVs in the absence of E5 proteins are still able to evade the immune system by interfering with IFNγ anti-viral pathway (Woodby et al., 2016). There are thus always compensatory evolutionary mechanisms developed over time to achieve the same biological end.

The Gamma-6 species lack the E6 ORF and so do HPV214 (Nobre et al., 2009b). The E6 protein is essential in binding of p53 tumour suppressor protein and cell cycle dysregulation (Wallace and Galloway, 2015). It appears however that this lost function in HPV214 may be compensated for

by its E7 protein as it has a LxCxE motif for pRB binding, which is lacked by the other novel types (Table 2.4). It has been shown elsewhere that HPV108, another member of the Gamma-6 species can induce dysplasia in organotypic keratinocytes without having the E6 protein (Nobre et al., 2009a), alluding to the fact that once the E6 binding function is not present there are compensatory mechanisms developed through evolution for the viruses to adapt to the host. It was recently shown that Gamma-6 HPVs acquired a 37 aa protein named E10, the E10 protein is upstream of the E7 start codon (Van Doorslaer and McBride, 2016, Van Doorslaer et al., 2017a) (Figure 2.2). In this study, E10 was identified in HPV214. However, the acquisition of E10 does not likely compensate for all E6 functions.

Zinc fingers are finger-like zinc binding domains in protein sequences. These are essential in protein to protein interactions and binding to DNA (Wayengera, 2012). The E6 and E7 proteins of the novel HPVs have Zinc finger binding domains. In HPV16, the zinc domains facilitate E6 binding to a number of cellular proteins including p53 (Wallace and Galloway, 2015). In high risk HPV types, the binding of p53 by E6 protein, in combination with the E7 protein binding of pRB, constitutes the hallmark of HPV carcinogenesis (Munger et al., 2004). The E7 protein of most oncogenic HPVs contain a pRB binding domain represented as LxCxE (Doorbar et al., 2015), and of the novel types only HPV214 and HPV222 contained this domain.

We found nuclear localisation like sequences (NLS) in E1, E2, L1 and L2 proteins of all the novel HPV types. In the L1 and L2 proteins they were NLS-like signals with slight modifications from the KRK and KRRL signatures, as reported in HPV16 (Zhou et al., 1991) and HPV199 (Ostrbenk et al., 2015). NLS signal motifs were identified in the C-terminus of the L1 and L2 proteins of the novel Gamma-HPVs with the motif (K/R)R(K/R), these have also been described elsewhere

(Zhou et al., 1991, Nelson et al., 2000). E1 proteins play a primary role in viral DNA replication and hence are found in the nucleus of the host cell. To facilitate this function, E1 proteins have amino acid sequences that are necessary for directing this nuclear localisation (Ostrbenk et al., 2015). E1 proteins have a bipartite NLS composed of two clusters separated by about 27-30 aa (KRK and KRRL) among the novel types, they were separated by 28aa. Another conserved NLS motif RKRxR/KRRR/KRXR, previously described in Alpha-HPVs (Zou et al., 2000) and also promotes nuclear localization was also found in the E2 proteins of the Gamma-HPVs.

Other domains identified include the PDZ binding domain in the N-terminal region of E6 (Bolatti et al., 2016), ATP binding site of E1, the DNA recognition helix of E2 and the furin cleavage site of L2. All these domains play a role in the HPV life cycle, the details of the mechanisms of their action are beyond the scope of this chapter. However, the transmembrane domain at the N-terminal of the L2 minor capsid protein has recently been described in HPV16 as essential for the translocation of viral DNA across phospholipid bilayer membranes (Bronnimann et al., 2013, Wang and Roden, 2013). The domain consists of $G(x)_3G$ motifs and similar glycine zippers $G(x)_3G(x)_3G$ motifs that together work in unison to facilitate packing of DNA helices to pass lipid bilayers. This domain was also identified in all the L2 proteins of the novel types but its functionality in Gamma-HPVs has not been explored.

LCR palindromic E2 binding sites (ACC-N_6-GGT) Li et al. (2009) were present in all the novel HPV types, these have also been described elsewhere (Newhouse and Silverstein, 2001). The origin of replication of PVs lies in the LCR and contains more than one E2 binding site (Sverdrup and Khan, 1995). The E2 protein acts as an activator as well as a repressor of viral transcription and initiation of replication, partitioning of genome and binds to two forms of palindromic sites:

ACC-$(N)_4$-CGGT and (ACC-N_6-GGT). The former site binds with higher affinity compared to the latter (Bedrosian and Bastia, 1990). The novel HPVs described in this study had the latter form of the palindromic binding sites which bind weakly, but this may not apply to the E2 from the novel types.

Polyadenylation sites are adenine rich and facilitate viral mRNA splicing. The early sites are positioned at the 5' end of the L2 protein, while the late polyadenylation sites are usually downstream of the L1 protein within the LCR (Chen et al., 2007a, Chen et al., 2007b). We identified early polyadenylation sites (-AATAAA-) at the N-terminal (5') end of the L2 protein of all the novel HPV types. The late polyadenylation sites were located in the URR region but at differing positions in each novel type.

In this study, we showed that HPV212, HPV213, HPV215, HPV220 and HPV222 novel types each had full genome clones with at least 99.5% pairwise identity to each other (Figure 2.9 and Figure 2.10). It has been suggested that differences in a single genetic region cannot be used to define a variant, but rather that the complete genome be used for variant classification instead of just the L1 ORF (Chen et al., 2015). A common nomenclature for HPV variants and sub-lineages using complete genomes has been implemented (Burk et al., 2013). The complete genomes of each of HPV212, HPV213, HPV215, HPV220 and HPV222 clones displayed <0.5% pairwise difference in nucleotide sequences. By strict definition a sub-lineage is an isolate of the same HPV type that differs from the other by a minimal of 0.5% and maximal of 1% difference (Calleja-Macias et al., 2005, Bernard et al., 2006, Burk et al., 2013) and hence the clones do not fit the definition of sub-lineages according to Burk and co-workers (Burk et al., 2013).

The intra-host genetic diversity of the novel Gamma-HPVs was examined here by Illumina sequencing of multiple whole genome clones of each type amplified from individual clinical specimens. This is the first study of intra-host variation of Gamma-HPVs. While there was no variation in the genomes of HPV211, HPV214, HPV216, HPV219 and HPV221, we identified 67, 51, 50, 17 and 24 variable nucleotide sites in the genomes of HPV212, HPV213, HPV215, HPV220 and HPV222 respectively. This diversity was greater than expected after having run two PCRs for each novel type to rule out polymerase-based artefacts. PVs have long been thought to have a low rate of mutation and to co-evolve with their hosts (Bravo and Felez-Sanchez, 2015, Chen et al., 2009, Dube Mandishora et al., 2018) due to the fact that they hijack the host cellular DNA replication machinery for replication which includes high fidelity polymerases with proofreading activity (with an error rate of about 4.3×10^{-5} (Korona et al., 2011)) and post-replication repair. However, similar to our findings, several recent NGS-based studies report high intra-host variability in Alpha-HPVs in clinical specimens, the studies using Ion Torrent sequencing of long PCR amplicons of the HPV16 genome identified between 3 to 175 variable nucleotide sites per genome in samples (Dube Mandishora et al., 2018, de Oliveira et al., 2015, Hirose et al., 2018). Hirose and co-workers (Hirose et al., 2018) identified an average of 7 nucleotide variations (range 0 to 85) per sample in the genomes of HPV16, HPV52 and HPV58. Dube Mandishora and co-workers (Dube Mandishora et al., 2018) identified hundreds of variant sites in the PGMY region of the L1 gene of HPV16, HPV18 and HPV52 using Illumina sequencing. Several explanations have been proposed for the high variability observed, including the generation of variants during infection due to the activation of host polymerases with lower fidelity (Dube Mandishora et al., 2018) as well as the recruitment of members of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of

mutagenic enzymes (Hirose et al., 2018, Mirabello et al., 2017, Vartanian et al., 2008) resulting in mutations rates higher than the human autosomal mutation rate.

The d_N/d_S ratios were calculated, with a ratio of one indicating neutral selection, >1 diversifying positive selection and <1 negative or purifying selection (Chen et al., 2009). The d_N/d_S ratios for all ORFs of HPV212, HPV213, HPV215, HPV220 and HPV222 were less than 1 indicating purifying selection (Table 2.6 and Table 2.7). This is in agreement with previous findings that the ORFs of HPV16 and other Alpha-HPVs are under strong purifying selection (DeFilippis et al., 2002, Carvajal-Rodriguez, 2008, Chen et al., 2005). Purifying selection is likely the result of the requirement for maintaining functional viral proteins for the successful completion of the virus life cycle. Gamma-HPVs are usually commensals with non-pathogenic and non-oncogenic potential and hence the absence for the need to be constantly under diversifying positive selection as a result of the need for immune escape. The combination of the lack of diversifying positive selection coupled with high intra-host variation in 5 of the novel HPV types maybe indicative of the fact that the variation observed is not an evolutionary event but random changes during replication of viral genome. This is supported by the fact that if it were an evolutionary event, we would have also identified more variations occurring among conserved functional domains. There were no non-synonymous changes in conserved functional domains except for a Cysteine to Phenylalanine substitution at the last position of the E7 Zinc finger binding domain found in two HPV213 clones. Whether this amino acid change is deleterious to the functioning of E7 was not explored.

Strikingly, we also observed that all of the novel HPV types had dN/dS values of less than 1 (indicative of purifying selection) in both the E6 and E7 protein (Figure 2.11). In Alpha-HPVs, the

E6 and E7 play a major role in oncogenesis as they have the retinoblastoma binding domain and Zinc finger binding domains essential in binding the p53 tumour suppressor protein. The lack of diversifying selection and presence of selection in the E7 of most Gamma-HPVs are probably indicative of the non-oncogenic potential of these viruses, as they do not exhibit any selection pressure.

2.4 CONCLUSIONS

The characterisation and classification of HPV211, HPV212, HPV213, HPV214, HPV215, HPV216, HPV219, HPV220, HPV221 and HPV222 add these novel types to the repertoire of the ever expanding Gamma-HPV genus. We make the fourth announcement of an HPV lacking the E6 ORF. It is apparent from the phylogenetic tree (Figure 2.4) based on L1 nucleotide sequences that the four members of the Gamma-6 species, besides the unusual genome lacking the E6 ORF, also have a considerable phylogenetic distance from other HPV types. Further investigations into the Gamma-6 species, tissue tropism and potential disease association of the other novel HPVs described here is warranted, in order to empirically evaluate their clinical significance. We recommend further studies into intra-host viral diversity.

Chapter 3: Evolutionary Dynamics of Ten Novel Gamma-PVs: Insights from Phylogenetic Incongruence, Recombination and Phylodynamic Analyses

3.0 INTRODUCTION

The discovery of numerous new PVs using next generation sequencing methods has begun to shed further light on the evolutionary history of this virus family. However, unravelling the evolutionary history of these viruses is potentially complicated by both inter-gene phylogenetic incongruence and recombination (Van Doorslaer, 2013). It has been observed that the nucleotide and encoded amino acid sequences of the E and L genes have evolved slightly differently in terms of evolutionary rates and selection pressures (Harari et al., 2014, García-Vallvé et al., 2005), a factor that could contribute to phylogenetic incongruence between the E and L gene trees. As a consequence of this, no single gene tree will accurately and adequately represent the evolutionary history of complete PV genomes (Van Doorslaer, 2013). Recombination events between different PVs may provide an additional explanation for gene-to-gene phylogenetic incongruence.

The genetic diversity and plurality of PVs along with the high frequencies of observed HPV co-infections make it reasonable to hypothesize recombination (Angulo and Carvajal-Rodriguez, 2007). However, technical difficulties associated with the inaccurate alignment of highly diverse PV gene sequences has slowed the study of recombination among PVs (Posada and Crandall, 2001). One of the most cited methods for recombination detection is the RDP program, which is a combination of rigorous alignment quality testing and powerful recombination detection methods (Varsani et al., 2006). During recombination detection, RDP4 rigorously tests the quality of sequence alignments to guard against the detection of false-positive recombination signals that arise due to sequence misalignment (Varsani et al., 2006).

We report in this chapter phylogenetic incongruences, recombination analysis and time evolution of the ten novel Gamma-PV types from humans, HPV211-HPV216 and HPV219-HPV222 (Murahwa et al., 2018), which we use along with all currently known members of the Gamma-PV genus and other PVs to describe the evolution of PVs.

3.1 METHODS

3.1.1 SOURCE OF SEQUENCE DATA

Sequences were downloaded from PaVE database (https://pave.niaid.nih.gov accessed on 27/01/2018). The sequences for HPV211-HPV216 and HPV219-HPV222 were obtained from our group (Murahwa et al., 2018). PVs are classified by the International Committee on Taxonomy of Viruses (Van Doorslaer et al.) into distinct species but the nomenclature of types can be done by specific working groups. Some of the viruses used in this study are pending classification by the ICTV, but were provisionally grouped into specific genera and types by the HPV Reference Centre in Sweden (http://www.nordicehealth.se/hpvcenter/reference_clones/).

3.1.2 PHYLOGENETIC TREE CONSTRUCTION

Construction of phylogenetic trees for incongruence tests: The sequences of the ten novel HPV types and the ten publicly available HPV sequences that were most closely related to each of these were used to construct phylogenetic trees. Phylogenetic trees were constructed based on MUSCLE (Edgar, 2004a) alignments of the nucleotide sequences of each ORF obtained from the PAVE database (Van Doorslaer et al., 2013a). Maximum likelihood trees were generated with PhyML 3.0 (Guindon et al., 2010) using the general time reversible nucleotide substitution model with discrete gamma rate heterogeneity among sites and invariable sites (GTR+G+I), as determined to be the best fitting nucleotide substitution model by jmodeltest (Darriba et al.,

2012). The approximate likelihood ratio test (aLRT) was used to estimate branch support (Anisimova and Gascuel, 2006). The trees were visualised in iTOL (http://itol.embl.de/upload.cgi) (Letunic and Bork, 2016).

3.1.3 PHYLOGENETIC INCONGRUENCE TESTS

We applied two different phylogenetic incongruence tests the Ktreedist test as a preliminary test and the Shimodaira-Hasegawa test.

The Ktreedist test: To determine the joint differences in topology and branch lengths between phylogenetic trees constructed from different HPV genes (E1, E2, E4, E7, L1, and L2), we utilised Ktreedist v1.0 (Soria-Carrasco et al., 2007). In brief, the Ktreedist program first scales a comparison tree to match the global divergence of a reference tree and then measures minimum branch length distances (BLD) between the trees: known as the tree K-score. K-scores close to 0 indicate that the phylogenetic trees being compared are very similar, i.e. they have similar branching orders and relative branch lengths, irrespective of their global evolutionary rates (Soria-Carrasco et al., 2007). High K-scores imply that trees are incongruent (Soria-Carrasco et al., 2007).

The Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999) **using W-IQ-TREE** (Trifinopoulos et al., 2016)**:** We used clustal alignments (Sievers et al., 2011, Li et al., 2015, McWilliam et al., 2013) of E1, E2, E4, E7, L1 and L2 HPV genes from 80 Gamma-HPVs (including the ten novel types) to compute the log-likelihoods of phylogenetic trees in W-IQ-TREE, which is a fast online phylogenetic tool for maximum likelihood analysis (http://iqtree.cibiv.univie.ac.at) (Trifinopoulos et al., 2016). The tool test tree topology estimates model parameters such as substitution rates and optimizes tree branch lengths to lessen computational usage. We used

default settings of the W-IQ-Tree, including best fit model (Kalyaanamoorthy et al., 2017) and ultra-fast bootstrap analysis (1000 alignments) (Minh et al., 2013) to run tree topology analysis including the Kishino-Hasegawa (KH) test (Kishino and Hasegawa, 1989), Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) and approximately unbiased (AU) test (Shimodaira, 2002) to test if there is a difference in evolutionary patterns amongst the different HPV genes. The results of the Shimodaira-Hasegawa test using W-IQ-Tree are shown in the appendix to this chapter as supporting evidence.

3.1.4 RECOMBINATION ANALYSIS

Seventy complete Gamma-PV genomes that are representative of all currently known Gamma-PVs obtained from the PAVE database, were combined into a dataset of 80 sequences with the ten novel HPV types recently discovered and genomically characterised by our group (Murahwa et al., 2018). All genomes were linearized at the first nucleotide position of their E6 genes except for Gamma-6 species viruses that lack E6 for which the start was shifted to the first nucleotide of their E7 genes. We then constructed an alignment containing the 80 Gamma-PVs using MUSCLE. This alignment was analysed using RDP v4.95 (Martin and Rybicki, 2000) with default settings, which implements analysis of recombination using several methods: RDP (Martin and Rybicki, 2000), BOOTSCAN (Martin et al., 2005a), CHIMAERA (Martin et al., 2005b), GENECONV (Padidam et al., 1999), MAXIMUM X^2 (Smith, 1992) and SISCAN (Gibbs et al., 2000).

3.1.5 CONSTRUCTION OF A TIME-SCALED HPV PHYLOGENY

We sought to infer the probable divergence times of our newly characterised HPV types from currently known HPVs. The complete L1 nucleotide sequences from 214 PV sequences were selected for analysis (Table 3.1). Two avian PVs: FcPV (*Fringilla coelebs*, the common chaffinch),

PePV (*Psittacus erithacus*, the grey parrot) and one turtle PV: CcPV1 (*Caretta caretta*, the loggerhead turtle) were also included in the analysis as outgroups. We performed a Bayesian evolutionary molecular clock analysis using BEAST v1.8.4 (Drummond et al., 2012). With a GTR+I+G nucleotide substitution model and an uncorrelated lognormal relaxed clock model. A fixed mean substitution rate for HPVs was applied based on estimated evolutionary rates inferred from a study that investigated the times to the MRCA based of Feline papillomaviruses (1.95×10^{-8} nucleotide substitutions per site per year) (Rector et al., 2007). The Markov Chain Monte Carlo (MCMC) analysis was run for 100,000 million generations with sampling every 10,000 generations. The final MCMC sampling chains were visually assessed for convergence and good mixing using Tracer v1.7.1 (Rambaut et al., 2018). A Maximum Clade Credibility (MCC) tree was generated after discarding the first 1000 trees that were obtained prior to the burn-in period of the chains.

Table 3.1 Summary of Analysis done and sequence dataset used.

Type of analysis	**Sequence data set used**
Phylogenetic tree construction	80 whole genomes of currently known Gamma-HPVs
K-treedist test (run as a preliminary test for incongruence)	10 novel Gamma-HPV types and 10 closely related Gamma-HPV types (also used for construction of preliminary gene phylogenetic trees in Figure 3.2)
Shimodaira-Hasegawa test	80 whole genomes of currently known Gamma-HPVs
Recombination analysis	80 whole genomes of currently known Gamma-HPVs
Time scaled HPV phylogeny	214 L1 nucleotide sequences of mostly HPV sequences, two avian PVs and one turtle PV.

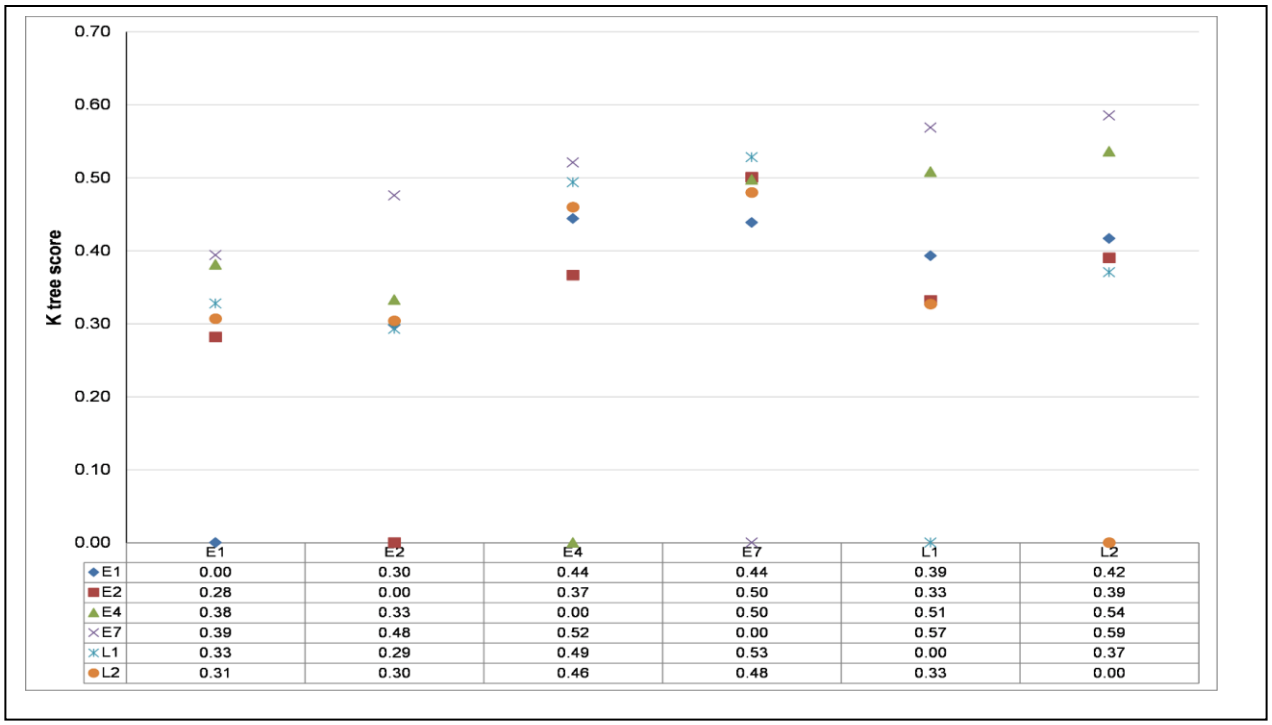
3.2 RESULTS

3.2.1 PHYLOGENETIC INCONGRUENCE AMONG NOVEL GAMMA-PVS GENE TREES

The Ktreedist yielded K-scores >0.45 which are indicative of substantial topological and branch-length differences between many different pairs of trees. This was especially true for the early and late gene comparisons (Figure 3.1). The highest degree of incongruence was between L1 and E7 (a K-score = 0.585; Figure 3.2). A high K-score (0.5) was also observed for the E7 vs E2 gene comparisons. Conversely, the L1 and L2 gene comparisons yielded the lowest K-scores (0.33). K score values are not symmetrical i.e. L1:L2 K score value differed from L2:L1 K score value. This preliminary result done for the 10 novel types and 10 of the closest known HPVs prompted further investigation of incongruence trees among all the currently known 80 Gamma-HPVs.

To determine if the phylogenetic trees for different Gamma-PV genes were congruent we then used a more conclusive test, the SH test (Shimodaira and Hasegawa, 1999). The null hypothesis of the SH test states that the difference between trees (branch length, topology or likelihoods) is zero. The observed differences were significantly greater than zero and rejected the null hypothesis for most of the gene versus gene comparisons, but particularly for the E gene versus L gene comparisons. Table 3.2 shows the results of the SH test using W-IQ-Tree, indicating that there is substantial phylogenetic incongruence between the late and early genes of Gamma-PVs as shown by the p-values (indicated in light blue in Table 3.2). The observed differences (deltaL values) were significantly greater than zero as shown by the p-values and allows us to reject the null hypothesis and declare that the trees are significantly different i.e. incongruent ($p < 0.05$).

The W-IQ-Tree results output are displayed in the appendix to this chapter (Supplementary Figures 3.6).



	E1	E2	E4	E7	L1	L2
E1	0.00	0.30	0.44	0.44	0.39	0.42
E2	0.28	0.00	0.37	0.50	0.33	0.39
E4	0.38	0.33	0.00	0.50	0.51	0.54
E7	0.39	0.48	0.52	0.00	0.57	0.59
L1	0.33	0.29	0.49	0.53	0.00	0.37
L2	0.31	0.30	0.46	0.48	0.33	0.00

Figure 3. 1 Minimum branch length distance (K tree score) between phylogenetic trees constructed using different HPV genes.

K-score values of closer to zero indicate that the phylogenetic trees are very similar, i.e. have similar topology and relative branch lengths, irrespective of their global evolutionary rates [33]. A high K-score value is an indication of incongruence regardless of global evolutionary rates**.**

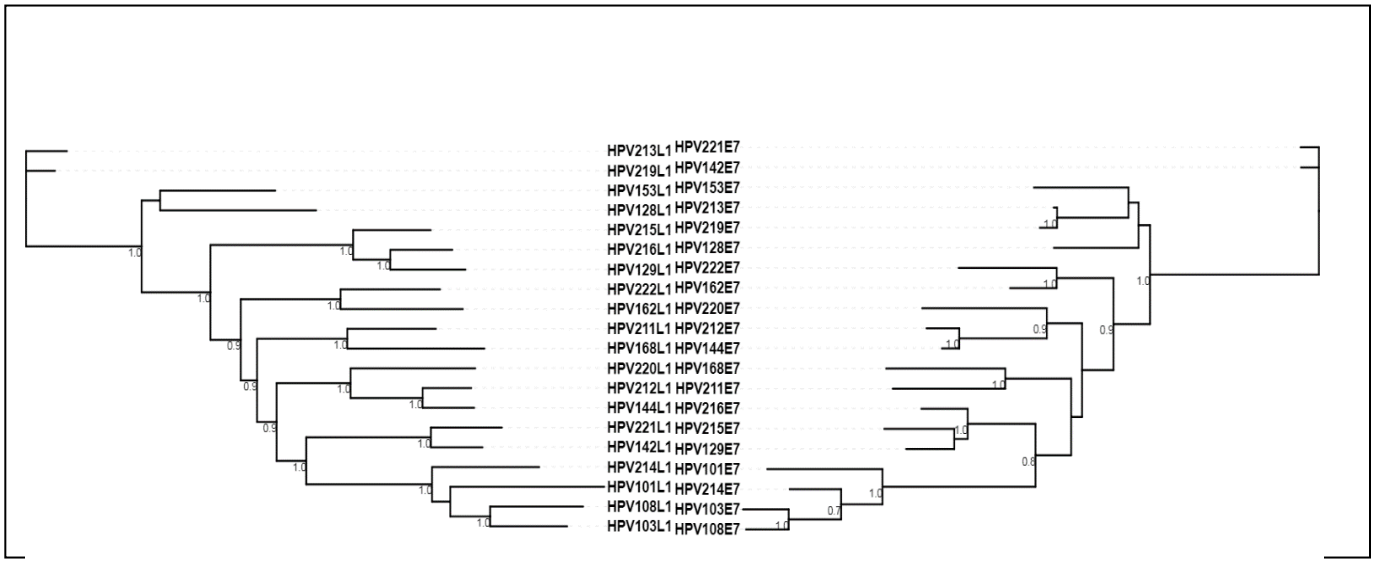


Figure 3. 2 A representative phylogenetic incongruence between early and late genes among the 10 novel types and their closest known Gamma-PVs.

Phylogenetic trees were inferred based on the L1 (left tree) and E7 (right tree) using MUSCLE alignment of the nucleotide sequences. Maximum likelihood trees were generated in PHYML with GTR substitution model with 1000 bootstrap replicates used for branch support estimation and viewed and edited in iTOL. Only the novel types were included in this analysis for readability. The dotted lines were added for ease of interpretation.

Table 3.2 Shimodaira-Hasegawa test for incongruence.

E1 as reference tree								
Tree	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
E1	0	1	1	1	1	1	1	0.999
E2	182.16	0	0	0.008	0	0	7.3e-33	0.000747
E4	639.14	0	0	0	0	0	3.8e-203	0.000845
E7	1048.70	0	0	0	0	0	0	1.8e-58
L1	381.21	0	0	0	0	0	3.8e-104	4.2e-05
L2	219.47	0	0	0.002	0	0	1.2e-40	2.4e-09
E2 as reference tree								
Tree	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
E1	137.43	0	0	0.023	0	0	1.3e-27	8.3e-05
E2	0	1	1	1	1	1	1	1
E4	288.35	0	0	0	0	0	1.3e-75	2.8e-07
E7	904.92	0	0	0	0	0	3.4e-299	6.7e-97
L1	273.21	0	0	0	0	0	5.9e-70	5.7e-44
L2	224.96	0	0	0	0	0	1.8e-60	5.1e-07
E4 as reference tree								
Tree	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
E1	79.74	0	0	0.012	0	0.004	0.000433	0.0217
E2	64.99	0.02	0.003	0.02	0.003	0.016	0.00211	0.104
E4	0	0.998	0.997	1	0.997	1	0.997	0.992
E7	311.86	0	0	0	0	0	8.7e-85	5.8e-114
L1	130.15	0	0	0	0	0	1.2e-23	6.4e-35
L2	84.83	0	0	0.004	0	0	1.1e-10	0.00483
E7 as reference tree								
Tree	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
E1	59374	0.004	0.016	0.017	0.016	0.064	0.00429	0.0117
E2	71.02	0.001	0.007	0.007	0.007	0.031	0.000919	0.00182
E4	138.43	0	0	0	0	0	1.78e-20	0.014
E7	0	0.946	0.952	1	0.952	0.991	0.944	0.945
L1	39.42	0.0041	0.048	0.119	0.048	0.189	0.043	0.118
L2	59.86	0.008	0.022	0.025	0.022	0.063	0.00759	0.0321
L1 as reference tree								
Tree	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
E1	188.69	0	0	0	0	0	1.3e-25	7.6e-05
E2	218.51	0	0	0	0	0	4.1e-40	6.6e-48
E4	473.21	0	0	0	0	0	1.6e-133	0.000267
E7	899.66	0	0	0	0	0	3.5e-301	5.98e-65
L1	0	1	1	1	1	1	1	1
L2	149.83	0	0	0.005	0	0	9.9e-301	1.23e-10
L2 as reference tree								
Tree	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
E1	226.34	0	0	0	0	0	2.4e-51	1.3e-52
E2	220.29	0	0	0	0	0	1.8e-50	8.4e-12
E4	478.78	0	0	0	0	0	1.5e-149	4.2e-06
E7	1087.1	0	0	0	0	0	0	0.00124
L1	229.72	0	0	0	0	0	1.3e-42	2.01e-13
L2	0	1	1	1	1	1	1	1

deltaL: logL difference from the maximal logl in the set.

bp-RELL: bootstrap proportion using RELL method (Kishino et al., 1990).

p-KH: p-value of one sided (Kishino and Hasegawa, 1989).

p-SH: p-value of Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999).

p-WKH: p-value of weighted KH test.

p-WSH: p-value of weighted SH test.

c-ELW: Expected Likelihood Weight (Strimmer and Rambaut, 2002).

p-AU: p-value of approximately unbiased (AU) test (Shimodaira, 2002).

3.2.2 RECOMBINATION ANALYSIS

The Gamma-PV whole genome alignments contained a total of only three plausible recombination events, namely event 1, 2 and 3. These events were all detectable by two or more different recombination detection methods with a p-value cut-off <0.05. However, event 1 and event 2 had no phylogenetic support and were therefore disregarded. Event 3 suggested that HPV4 and its near relatives all share evidence of the same ancestral recombination event (Figure 3.3 and Table 3.3). Only one of our newly discovered novel HPV types, HPV221, is a potential recombinant. The remainder of the potential recombination signals from event 3 are shown in Table 3.3.

Table 3.3 Inter-species recombination event 3 in whole genome Gamma-PVs.

Estimated Break point positions				Recombinant (Gamma-species)	Parent sequence		Evidence (method with P value < 0.05)	Intra-species or inter species recombination
In Alignment		In Genbank sequence			Minor (Gamma-species)	Major (Gamma-species)		
4446	8018	3631	6426	^HPV4 **(γ-1)**	HPV130 **(γ-10)**	Unknown (HPV162)**(γ-19)**	M,C,S,3S	Inter-species
				HPV163[T]				
				HPV173[T]				
				HPV205[T]				
				HPV158[T]				
				HPV95[T]				
				HPV65[T]				

N.B. Only event 3 is shown in Table 3.3, all the potential recombinant Gamma-PVs for this event are shown in the recombinant column and the proposed major and minor parents.

Abbreviations: M=MAXIMUM X^2, S= SISCAN, C=CHIMAERA, 3S=3SEQ

^ = The recombinant sequence may have been misidentified (one of the identified parents might be the recombinant.

Minor Parent = Parent contributing the smaller fraction of sequence.

Major Parent = Parent contributing the larger fraction of sequence.

Unknown = only one parent and a recombinant need be in the alignment for a recombination event to be detectable the sequence listed as unknown was used to infer the existence of a missing parental sequence.

[T] Sequences with trace evidence of the same recombinant event.

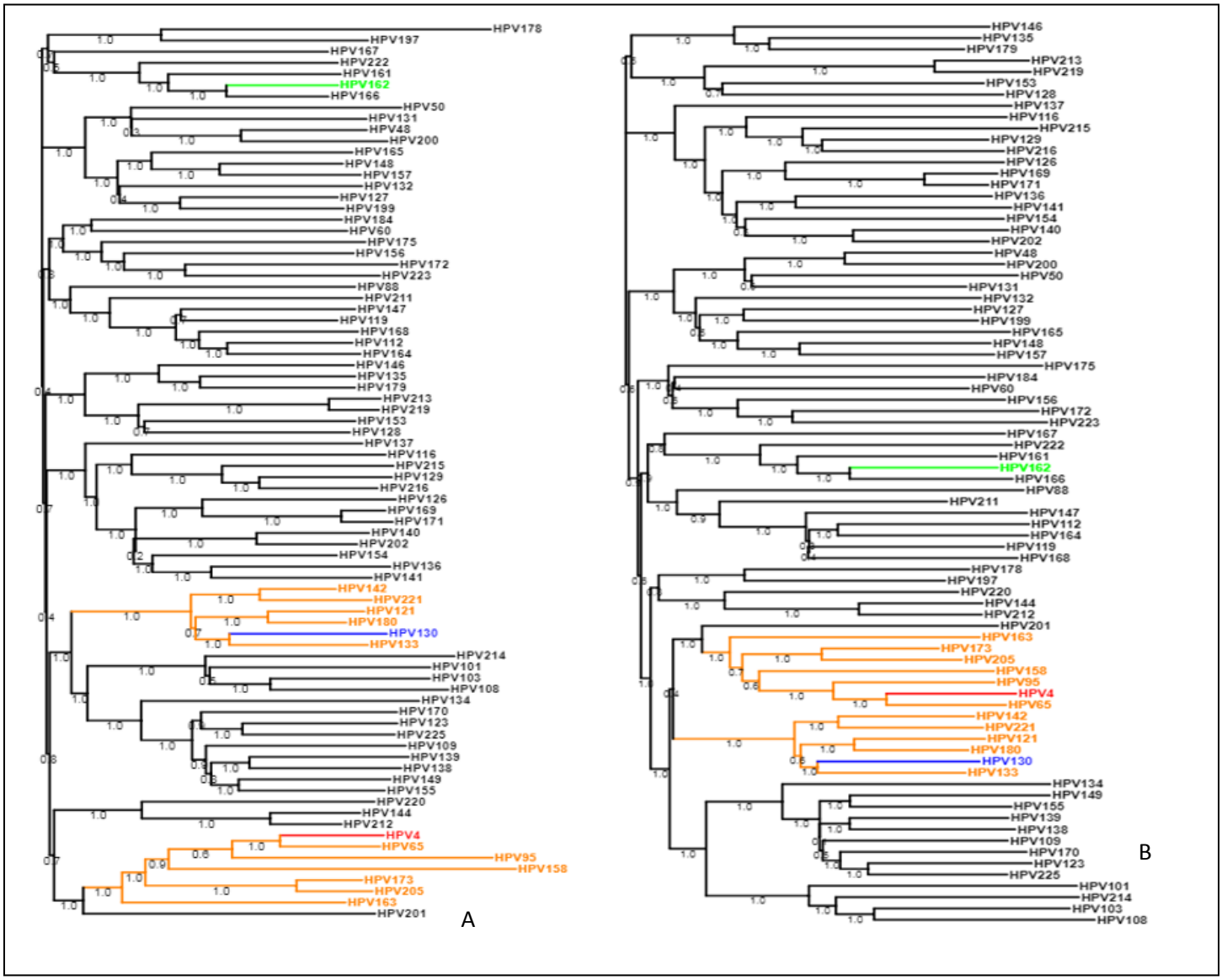


Figure 3. 3 Fast NJ tree from A) major parent (1-4445 and 8019-9792) and from B) minor parent (4446-8018).

Red-potential recombinant, Orange-sequence with trace evidence of the same recombination event, Blue-potential minor parent, Green-sequence used to infer unknown parent

Phylogenetic evidence of an ancient recombination event amongst Gamma-PVs. These unrooted neighbour joining trees (1000 bootstrap replicates, Jukes Cantor distances) were constructed in RDP v4.95 exported as Newick format and viewed in iTOL. Virus recombination methods indicate with a high degree of significance that the sequences in brown/orange are recombinant descendants of sequences in blue and green.

3.2.3 THE TIME-SCALE OF GAMMA-PV EVOLUTION

A fixed mean substitution rate for HPVs was applied based on estimated evolutionary rates inferred from that of Feline PVs (1.95×10^{-8} nucleotide substitutions per site per year) (Rector et al., 2007). The divergence times of the MRCA of HPV was predicted to have occurred 53.9 MYA (95% HPD 49.7-58.5), before splitting into the five main potential ancestors (Alpha, Beta, Mu, Nu and Gamma genera). The MRCA of the present day Gamma-PVs was predicted to have occurred approximately 49.8MYA (95% HPD 45.3-67.5). The novel HPV212 was predicted to have diverged from its closest relative, HPV144, about 7.6 MYA (95% HPD 5.2-10.4), the remaining nine novel Gamma-PVs divergence times from the MRCA are shown in Table 3.4. The predictions lie between 7.6 to 19.9 MYA.

Table 3.4 Mean divergence time of 10 novel HPVs from other gamma species or closest relative.

HPV types	Gamma-Species	Posterior	Mean Divergence Time (95% HPD) from MRCA
HPV211	8	1	19.9 (16.8 - 23.1) MYA from all other Gamma-8 species HPV types
HPV212	17	1	7.6 (5.2 - 10.4) MYA from its closest relative HPV144
HPV213	13	1	5.4 (4.0-7.1) MYA from its closest relative HPV219
HPV214	6	1	17 (12.8-20.7) MYA from closest relatives HPV108 and HPV103
HPV215	9	1	11.3 (9.2-13.7) MYA from closest relative HPV216 and HPV129
HPV216	9	1	8.3 (6.2-10.6) MYA from closest relative HPV129
HPV219	13	1	5.4 (4.0-7.1) MYA from its closest relative HPV219
HPV220	17	1	19.9 (15.1-23.0) MYA from all other Gamma-17 species HPV types
HPV221	10	1	8.3 (5.8-11.5) MYA from closest relative HPV142
HPV222	19	1	19.2 (15.7-23.0) MYA from all other Gamma-19 species HPV types

Since it has been reported that between 10 and 20 MYA, several hominoid precursors lived in Africa, Europe and Asia (Andrews, 1992), the timing and spatiotemporal patterning of Neanderthal human precursors' disappearance has only been radiocarbon dated to a limit of

50,000 years ago (Higham et al., 2014). Our MRCA prediction suggests that all the current HPV species diverged from what are currently their nearest relatives before the origin of modern humans.

The MRCA of the Gamma-6 species was predicted to have existed 22 MYA (95% highest posterior density (HPD) 17.7-27.3), earlier than that of most other Gamma-PVs species. The MRCA of all other Gamma-species viruses and the Gamma-6 viruses was predicted to have occurred 46.8 MYA (95% HPD 42.9-51.3). The rest of the node divergence times are shown in Figure 3.4 with the 95% highest posterior densities and the posterior support values are shown in Supplementary Figure 3.5 in the appendix to chapter 3.

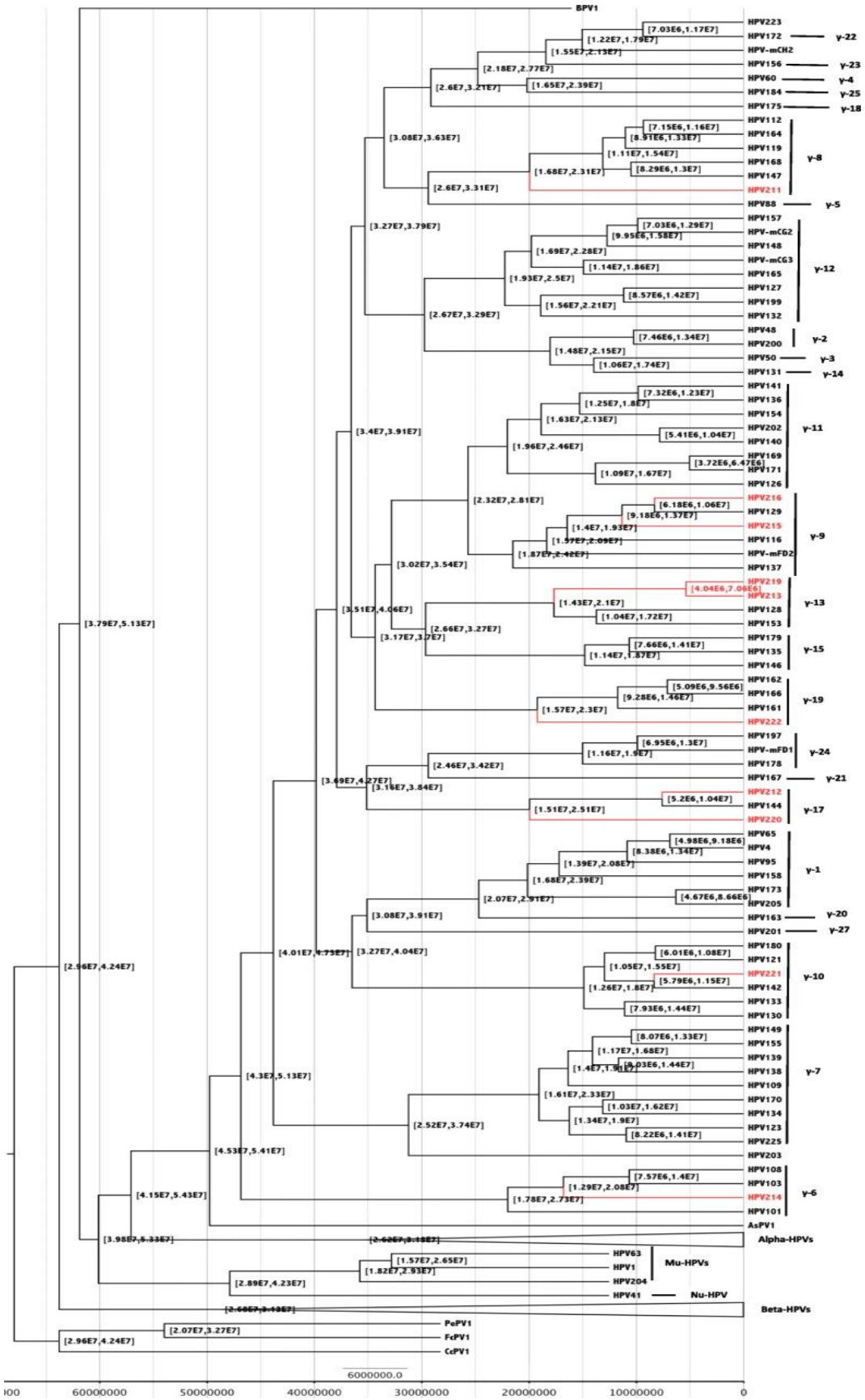

BPV1
HPV223
HPV172
HPV-mCH2
HPV156
HPV60
HPV184
HPV175
HPV112
HPV164
HPV119
HPV168
HPV147
HPV211
HPV88
HPV157
HPV-mCG2
HPV148
HPV-mCG3
HPV165
HPV127
HPV199
HPV132
HPV48
HPV200
HPV50
HPV131
HPV141
HPV136
HPV154
HPV202
HPV140
HPV169
HPV171
HPV126
HPV216
HPV129
HPV215
HPV116
HPV-mFD2
HPV137
HPV219
HPV213
HPV128
HPV153
HPV179
HPV135
HPV146
HPV162
HPV166
HPV161
HPV222
HPV197
HPV-mFD1
HPV178
HPV167
HPV212
HPV144
HPV220
HPV65
HPV4
HPV95
HPV158
HPV173
HPV205
HPV163
HPV201
HPV180
HPV121
HPV221
HPV142
HPV133
HPV130
HPV149
HPV155
HPV139
HPV138
HPV109
HPV170
HPV134
HPV123
HPV225
HPV203
HPV108
HPV103
HPV214
HPV101
AsPV1
HPV63
HPV1
HPV204
HPV41
PePV1
FcPV1
CcPV1
Alpha-HPVs
Mu-HPVs
Nu-HPV
Beta-HPVs
γ-22
γ-23
γ-4
γ-25
γ-18
γ-8
γ-5
γ-12
γ-2
γ-3
γ-14
γ-11
γ-9
γ-13
γ-15
γ-19
γ-24
γ-21
γ-17
γ-1
γ-20
γ-27
γ-10
γ-7
γ-6
[7.03E6,1.17E7]
[1.22E7,1.79E7]
[1.55E7,2.13E7]
[2.18E7,2.77E7]
[2.6E7,3.21E7]
[1.65E7,2.39E7]
[3.08E7,3.63E7]
[7.15E6,1.16E7]
[8.91E6,1.33E7]
[1.11E7,1.54E7]
[1.68E7,2.31E7]
[8.29E6,1.3E7]
[2.6E7,3.31E7]
[3.27E7,3.79E7]
[7.03E6,1.29E7]
[9.95E6,1.58E7]
[1.69E7,2.28E7]
[1.14E7,1.86E7]
[1.93E7,2.5E7]
[8.57E6,1.42E7]
[1.56E7,2.21E7]
[2.67E7,3.29E7]
[7.46E6,1.34E7]
[1.48E7,2.15E7]
[1.06E7,1.74E7]
[3.4E7,3.91E7]
[7.32E6,1.23E7]
[1.25E7,1.8E7]
[1.63E7,2.13E7]
[5.41E6,1.04E7]
[1.96E7,2.46E7]
[3.72E6,6.47E6]
[1.09E7,1.67E7]
[2.32E7,2.81E7]
[6.18E6,1.06E7]
[9.18E6,1.37E7]
[1.4E7,1.93E7]
[1.57E7,2.09E7]
[1.87E7,2.42E7]
[3.02E7,3.54E7]
[4.04E6,7.06E6]
[1.43E7,2.1E7]
[1.04E7,1.72E7]
[3.51E7,4.06E7]
[2.66E7,3.27E7]
[3.17E7,3.7E7]
[7.66E6,1.41E7]
[1.14E7,1.87E7]
[5.09E6,9.56E6]
[9.28E6,1.46E7]
[1.57E7,2.3E7]
[6.95E6,1.3E7]
[1.16E7,1.9E7]
[2.46E7,3.42E7]
[3.69E7,4.27E7]
[3.16E7,3.84E7]
[5.2E6,1.04E7]
[1.51E7,2.51E7]
[4.98E6,9.18E6]
[8.38E6,1.34E7]
[1.39E7,2.08E7]
[1.68E7,2.39E7]
[2.07E7,2.91E7]
[4.67E6,8.66E6]
[3.08E7,3.91E7]
[3.79E7,5.13E7]
[4.01E7,4.75E7]
[3.27E7,4.04E7]
[6.01E6,1.08E7]
[1.05E7,1.55E7]
[5.79E6,1.15E7]
[1.26E7,1.8E7]
[7.93E6,1.44E7]
[2.56E7,4.24E7]
[8.07E6,1.33E7]
[1.17E7,1.68E7]
[9.03E6,1.44E7]
[1.61E7,2.33E7]
[1.03E7,1.62E7]
[1.34E7,1.9E7]
[8.22E6,1.41E7]
[4.3E7,5.13E7]
[2.52E7,3.74E7]
[4.53E7,5.41E7]
[7.57E6,1.4E7]
[1.29E7,2.08E7]
[1.78E7,2.73E7]
[4.15E7,5.43E7]
[3.98E7,5.33E7]
[2.62E7,3.18E7]
[1.57E7,2.65E7]
[1.82E7,2.93E7]
[2.89E7,4.23E7]
[2.68E7,3.19E7]
[2.07E7,3.27E7]
[2.96E7,4.24E7]
6000000.0
60000000
50000000
40000000
30000000
20000000
10000000
0

Figure 3. 4 Molecular divergence times of PVs.

Classification was based on (Bernard et al., 2010, de Villiers et al., 2004). Exact divergence estimates in million years (for the nodes corresponding to the 10 HPV types 95% highest posterior density are presented in Table 3.4). The novel types are indicated in red. Posterior support values are provided in the Supplementary Figure 3.5.

3.3 DISCUSSION

An analysis of phylogenetic trees generated from different genes of the ten novel South African Gamma-PVs and their closest known relatives indicated differences, ranging from slight to major, between the branch lengths and branching orders of phylogenetic trees constructed from different genes. The Ktreedist method was used as a preliminary test to quantify these differences. After confirming existence of incongruences a more rigorous test, the SH test using W-IQ-TREE was done. In W-IQ-TREE 80 known HPV sequences were used to construct gene phylogenetic trees that were then computed into the online tool to test for incongruence. Apart from SH test the W-IQ-TREE tool also tests for incongruence using other algorithms: the bp-RELL, p-KH, weighted KH and weighted SH, the c-ELW and the AU test (see legend to Table 3.2 and also supplementary figures 3.6). The SH test consistent with the other tests shows p-values of less than 0.05 indicative of differences among phylogenetic trees from different genes. To visualise the incongruence an example of 20 HPV types (the ten novel and their closest known HPV types) were used to construct an L1 and E7 trees as shown in Figure 3.2, and juxtaposed next to each other these trees show slightly different branching patterns.

Differences between internal branches of phylogenetic trees constructed using different parts of the same HPV genomes have been described extensively elsewhere (Bravo and Felez-Sanchez, 2015, Bravo and Alonso, 2004b, Schiffman et al., 2005a). These differences could imply the occurrence of recombination, as has been previously reported (Shah et al., 2010).

We did not, however find strong evidence of recombination among Gamma-HPV types. From the 80 full Gamma-PVs (70 known HPV types and the ten novel types) included in the analysis, only one strongly supporting a recombination event was detected. Recombination in Gamma-PVs has been reported elsewhere (Bolatti et al., 2016). In that study, seven potential recombination events were reported using separate analyses of individual genes (rather than analysing full genomes). We detected few Gamma-PVs recombination events (Varsani et al., 2006) reported 529 potential recombination events, yet only ten were true events. The phenomenon is not exclusive to Gamma-PVs, in Alpha-PVs recombination events have been well described (Angulo and Carvajal-Rodriguez, 2007, Narechania et al., 2005a).

The difficulty in detecting recombination events in PVs relates to the technical difficulty of aligning highly divergent DNA sequences (Bravo and Felez-Sanchez, 2015, Varsani et al., 2006, Angulo and Carvajal-Rodriguez, 2007). Another factor is that most recombination detection methods are only designed to detect recombination events when one or both of the parental sequences have close relatives represented in the dataset being analysed (Van Doorslaer, 2013). Therefore, as more PVs are discovered, it becomes more likely that sequences closely related to the parents of recombinants will be included in recombination analyses. We also only included Gamma-PVs, in our recombination analysis dataset and our analysis was therefore only powered to detect for evidence of intra-genus recombination. However, this is probably not a major issue since the only convincing evidence of recombination in PVs that has so far been published has been between PVs in the same genus (Narechania et al., 2005a). In chapter 2, it is reported that the ten novel Gamma-PVs are under no positive selection pressure but rather purifying selection (dn/ds <1).

It has been shown elsewhere that different PV genes are under different selection pressures (Rocha Rodríguez et al. (2012) and also that different genes have different evolutionary rates ranging between $2x10^{-8}$ and $5x10^{-9}$ substitutions per site per year (Shah et al., 2010, Rector et al., 2007). Thus, no single gene tree will accurately represent the evolutionary history of entire PV genomes (Van Doorslaer, 2013).

Consequently, we report here phylogenetic tree incongruence with no evidence of recombination. It has been proposed that in such scenarios there is convergent as opposed to divergent evolution (Castoe et al., 2009). Convergent evolution can be defined as the independent evolution of similar features or characteristics in species of different lineages.

We show here that the MRCA of HPV was predicted to have occurred about 50-60 MYA. This is comparable to work done by Chen *et al.* (Chen et al., 2007b), and that the MRCA of the Gamma-PVs existed about 45-67 (49.8) MYA. Further, we were able to show that within the Gamma-PV genus, the Gamma-6 species split from all the other Gamma species about 43-51 (46.8) MYA, while Van Doorslaer and Mcbride (2016) and (Van Doorslaer and McBride, 2016) showed that Gamma-6 species last shared a common ancestor with other Gamma-PVs around 60 MYA. Further, we showed that the MRCA of the Gamma-6 species occurred about 22 MYA, which concurs with Van Doorslaer and Mcbride (Van Doorslaer and McBride, 2016), who reported that the Gamma-6 species MRCA existed 23.4 MYA. We therefore hypothesise that the Gamma-6 species lost the E6 gene between 20 and 60 MYA. This suggests that E6 was lost before the evolution of hominoid primates between 10-20 MYA (Andrews, 1992). This implies two things: 1) that viruses lacking E6 may infect Old World and New World monkeys, suggesting that it could be productive to hunt for these viruses in primates, and 2) that the E6-minus

viruses co-evolved with their hosts over a long enough period of time for us to have been able to isolate them from current humans (assuming PV species specificity). Chen *et al.* (Chen et al., 2007b) had previously predicted the loss of the E6 gene to have occurred about 15-30 MYA, the estimate was based only on the L1 ORF of nine divergent papillomaviruses. Here, we have estimated this date using 214 papillomavirus L1 nucleotide sequences from all the genera containing HPV sequences.

PVs lacking E6 have also been described in parrots (PePV1), donkeys (EaPV1) and bovines (Willemsen and Bravo, 2018). Presently, seven known human PVs of the Gamma-6 species lack the E6. The size of E6 (mean 253.5 nt) in the genomes of PVs infecting birds and turtles is about half the size of that in mammal-infecting PV genomes (mean 438 nt) (Willemsen and Bravo, 2018). The larger size of the mammalian PV E6 accommodates a second E6 Zinc finger binding motif domain (Zanier et al., 2013): a domain that was possibly a duplication of an ancestral E6 motif (Suarez and Trave, 2018). E6 mediated p53 degradation has been described as one of the hallmarks of HPV-mediated carcinogenesis, and the presence of this double motif may explain the increased likelihood of HPVs in causing cancer compared to PVs infecting birds and turtles.

Gamma-PVs lack an E5 gene. The E5 gene is located between the early genes and the late genes and is thought to have evolved originally from a non-coding region that was integrated between the early and the late genes of an ancestral sequence belonging to the Alpha-PV lineage (Willemsen and Bravo, 2018). Willemsen and Bravo also suggest that integration of E5 in this region promoted an adaptive radiation which yielded E6 and E7 proteins capable of degrading tumour suppressor proteins and facilitating carcinogenesis (Willemsen and Bravo, 2018). This is supported by the fact that 1) E6 and E7 proteins in Alpha-PVs (together with E5)

have greater onocogenic potential as classified by IARC as compared to the E6 and E7 of Gamma-PVs (that lack E5) and 2) E5 has an evolutionary rate that is approximately twice that of the remainder of the PV genome. The integration of the E5 ORF was predicted to have occurred in an ancestral virus that existed 30-60 MYA (in the Cenozoic era) which eventually gave rise to the Alpha, Mu and Nu lineages; each of which has different cell tropisms and clinical manifestations (Bravo and Felez-Sanchez, 2015). Willemsen et al. (Willemsen and Bravo, 2018) inferred the appearance of the E5 oncogene occurred 53-58 MYA, well within the range of that predicted by Bravo et al (Bravo and Felez-Sanchez, 2015).

We also reported the acquisition of E10 in HPV214, which we hypothesised coincided with E6 loss as previously reported (Van Doorslaer et al., 2017a). We speculate that if the loss of E6 occurred 20-60 MYA then E10 was acquired a few million years later or that the loss and gain might have occurred concurrently as a modification of E6 to E10. This is supported by the fact that the E10 ORF overlaps with the more conserved portion of the E6 scar (Van Doorslaer and McBride, 2016).

We report divergence times from 7.6 to 20 MYA with most lying well within the origin times of many other known PVs. However, HPV211 of the Gamma-8 species branches earlier from the other five HPV types in this species, i.e. the MRCA of HPV211 and the other 5 Gamma-8 species types is predicted to have occurred 20 MYA. This implies that HPV211 is closer to the ancestral or parental sequence of Gamma-8 species compared to HPV112, HPV119, HPV168, HPV147, and HPV164, and hence, it is more likely to be major/minor parent than it can be a recombinant. The MRCA of HPV222 and the other 3 members of the Gamma-19 species is 19.2 MYA. HPV222 branches from HPV161, HPV162 and HPV166 earlier than the others branch from

each other, also making it closer to the ancestral or parental sequence of the Gamma-19 species.

3.4 CONCLUSION

In this work, we report on the evolutionary characterisation of Gamma-PVs including that of the novel HPV types. To get a deeper insight into the evolutionary processes that may influence the diversification of Gamma-PVs, we explored phylogenetic incongruences among different genes of the novel types, attempted to discover potential recombination events between all known Gamma-PVs, and also estimated the time scale for Gamma-PV evolution. Consequently, we report here phylogenetic tree incongruence without strong evidence of recombination.

Chapter 4: General Discussion and conclusions

The aim of this thesis after the discovery of ten novel HPV viruses using NGS was to isolate the whole genomes of these ten viruses. This was achieved using back to back primers in a long-range PCR using a touch down approach (in terms of annealing temperature) to increase primer binding specificity. Once the whole genomes were amplified, they were cloned and then deep sequenced using Illumina MiSeq to obtain full genome sequences. The characterisation of the full genomes of the novel viruses is described in the second chapter of this thesis. The genomic organisation, phylogenetic clustering, genomic variation, identification and description of the conserved functional domains in the novel HPV types is also discussed in chapter 2. The genomes of all the novel viruses belonged to the Gamma-PVs genera but from different species of the genus. A phylogenetic tree of all known HPVs is illustrated in Figure 2.4 and shows the distribution of these viruses among other Gamma-HPVs. Most showed classical genomic organisation of the Gamma-HPVs save for HP214, which lacked the E6 ORF and additionally had a putative ORF, the E10. HPV212, HPV220 and HPV222 also had a premature stop codon in the E4 ORF and hence their principal E4 protein was in the form of E1^E4 transcript. Several conserved domains were identified in the novel HPV types. Among these domains were: the Zinc finger binding domain, PBM, NLS and NES, ATP binding domain, transmembrane binding domain, furin cleavage sites, DNA recognition site, polyadenylation sites, TATA binding box and the E1 and E2 binding sites of the LCR. Functional experiments were not done and hence this thesis did not attempt to describe the functionality of these domains but rather was limited to identification of these domains.

Of the 10 novel viruses discovered, half showed no intra-sample variation (HPV211, HPV214, HPV216, HPV219, HPV221) while the other half (HPV212, HPV213, HPV215, HPV220 and HPV222) had several "variants" within individual samples.

The intra-sample variation was however less than 0.5%. We ruled out PCR artefacts because two PCR runs were done for each potential novel HPV type, and Illumina sequencing has a very low error rate. The conclusion drawn from this is that HPVs evolve slowly and the low percentage variation is indicative of the robustness of host DNA polymerases. Also deduced from this chapter is the dn/ds ratios less than one, with a ratio of one indicating neutral selection, >1 diversifying positive selection and <1 negative or purifying selection (Chen et al., 2009). These novel Gamma-HPVs are certainly under no positive selection pressure, immune or vaccine induced. Their negative selection maybe suggestive of the fact that they are commensal organisms.

Chapter 3 focuses on the evolutionary dynamics of Gamma-HPVs with insights from phylogenetic incongruence, recombination and phylodynamic analysis (molecular divergence). For phylogenetic incongruence we focused only on the ten novel viruses and ten of their closest known HPV types in order to confirm what has been described for all known PVs: that different gene regions evolve differently (Harari et al., 2014, García-Vallvé et al., 2005). Once we confirmed phylogenetic incongruence of 20 Gamma-HPVs inclusive of the novel types, we then sought to establish these among all the 80 currently known and curated Gamma-HPV types. To this effect we did a SH test, which is more conclusive than the former. Having confirmed phylogenetic incongruences we then sought to investigate the existence of recombination events among the novel types and the known Gamma-HPVs. Information on HPV

recombination is scarce and varied depending on the method used. Hence, in the recombination analysis all 80 Gamma-PVs (10 novel types included) were analysed. It is important to note that by so doing we only explored intra-genus recombination, it has been suggested that inter-genus recombination may occur (Shah et al., 2010). Our recombination results actually demonstrated one recombination result which is not highly supported and is a Gamma-HPV inter-species recombination between HPV4 (Gamma-1) and HPV130 (Gamma-10) and HPV162 (Gamma-19). Detecting recombination in HPV is hampered by technical difficulties in alignment and there is scarce information on this subject despite the fact that there are about 220 HPV types and that a lot of multiple infections have been reported in several studies, which theoretically is good grounds to suspect recombination (Bolatti et al., 2016). The fact that we do not know the whole list of HPVs, offers the potential of more recombination events being described once the discovery efforts are continued and more HPV sequences are made available in the repositories. The improvement of alignment algorithms is also essential in refining the search for recombination events among HPVs. In the phylodynamic analysis of HPV, we used a tree inferred from L1 nucleotide sequences of 214 PVs and included some animal PVs as an outgroup: Two avian PVs; FcPV (*Fringilla coelebs*, the common chaffinch), PePV (*Psittacus erithacus*, the grey parrot) and one turtle PV: CcPV1 (*Caretta caretta*, the loggerhead turtle). The use of the L1 tree to infer phylogeny has been described elsewhere (Chen et al., 2007b). While it is generally known that different regions of the PV genome evolve differently, the L1 gene is the most highly conserved and slight changes in this region are likely to reflect on the divergence times of the overall genome. The fact that they were discovered in Africa which is believed to be the cradle of mankind prompted further study into the timescale of these viruses compared to human evolution. The divergence times of the MRCA of HPV in this thesis was

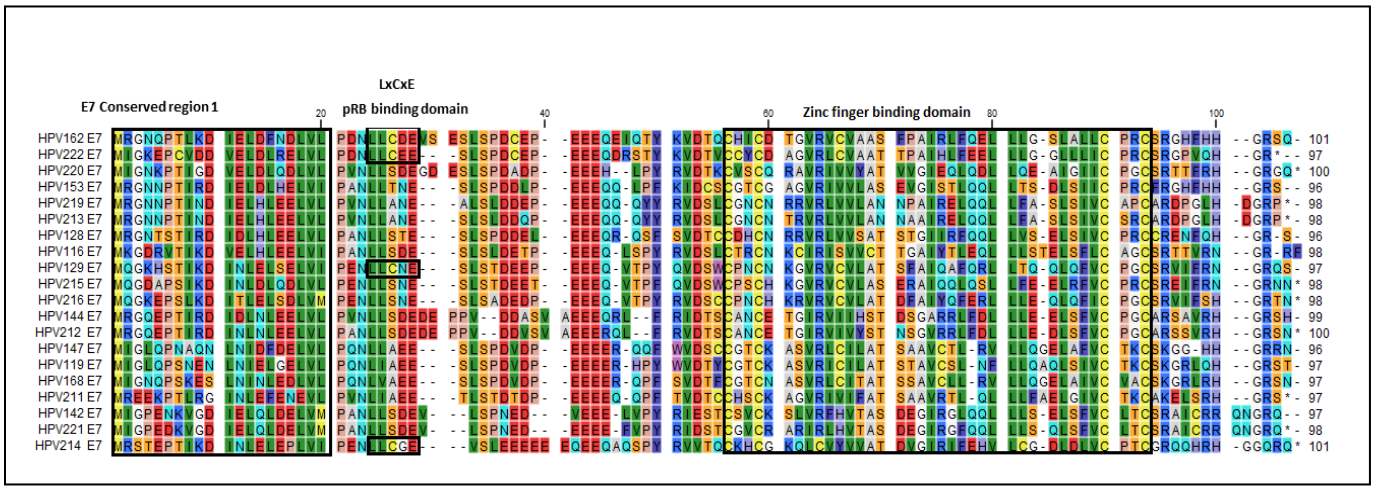
predicted to have occurred 53.9 MYA. Several hominoid precursors lived in Africa, Europe and Asia between 10 and 20 MYA (Andrews, 1992), the timing and spatiotemporal patterning of Neanderthal human precursors disappearance has only been radiocarbon dated to a limit of 50,000 years ago (Higham et al., 2014). Our prediction suggests that all the current HPV species diverged from what are currently their nearest relatives before the origin of humans.

In conclusion, ten novel Gamma-HPVs were discovered and characterised. Through various bioinformatics means we enriched our knowledge of this genus and other HPVs in general by including and comparing the novel types with already known HPV types.

Appendix 1: Appendix to Chapter 2

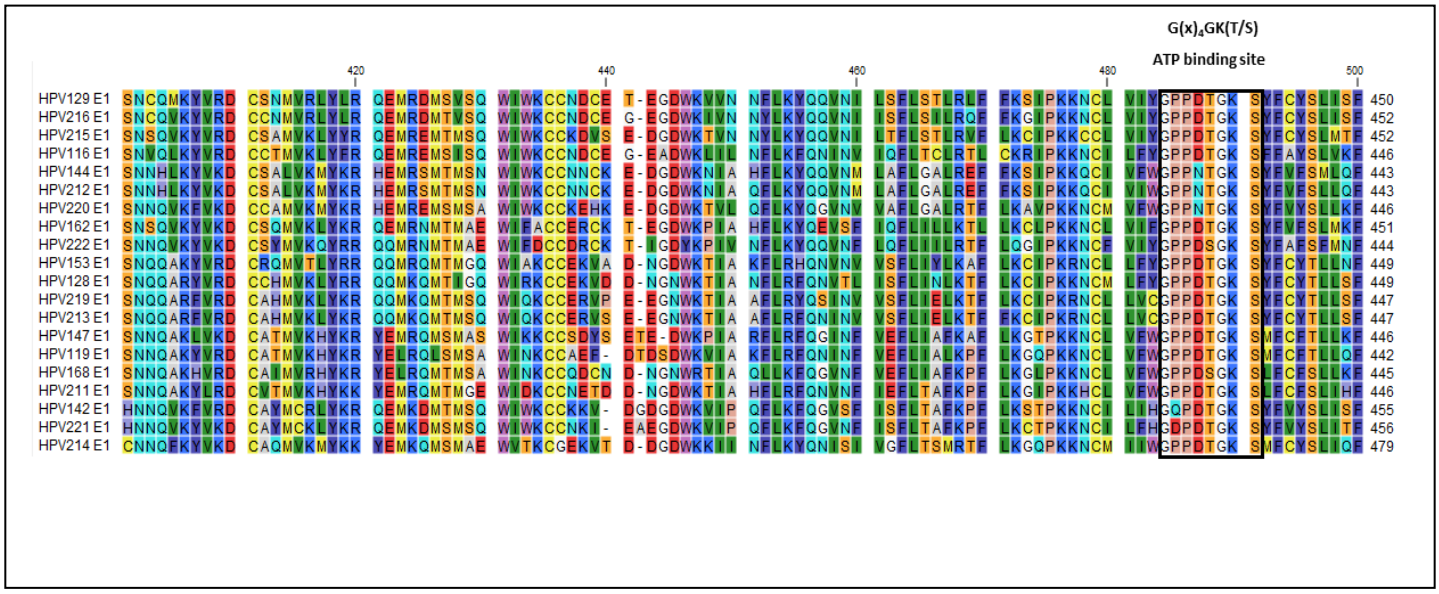
Supplementary Table 2.9 E1^E4 and E2^E8 spliced gene product prediction checklist.

	CHECK POINTS
1	E1^E4 ATG is same as E1 ATG
2	E8^E2 ATG is in +1 frame of E1
3	E1 portion of E1^E4 ends with AG and spliced portion starts with GT within E2
4	E8 portion of E8^E2 ends with AG and spliced portion starts with GT within E2
5	Start of E2 in E8^E2 and E4 in E1^E4 is same
6	There is an AG before start of E2 in E8^E2 and E4 in E1^E4
7	E1^E4 stop codon = E4 stop codon
8	E8^E2 stop codon = E2 stop codon
9	E2 portion of E8^E2 is identical to C-terminus E2
10	E4 portion of E1^E4 is identical to E4
11	E4 portion of the E1^E4 is in +1 frame of E2



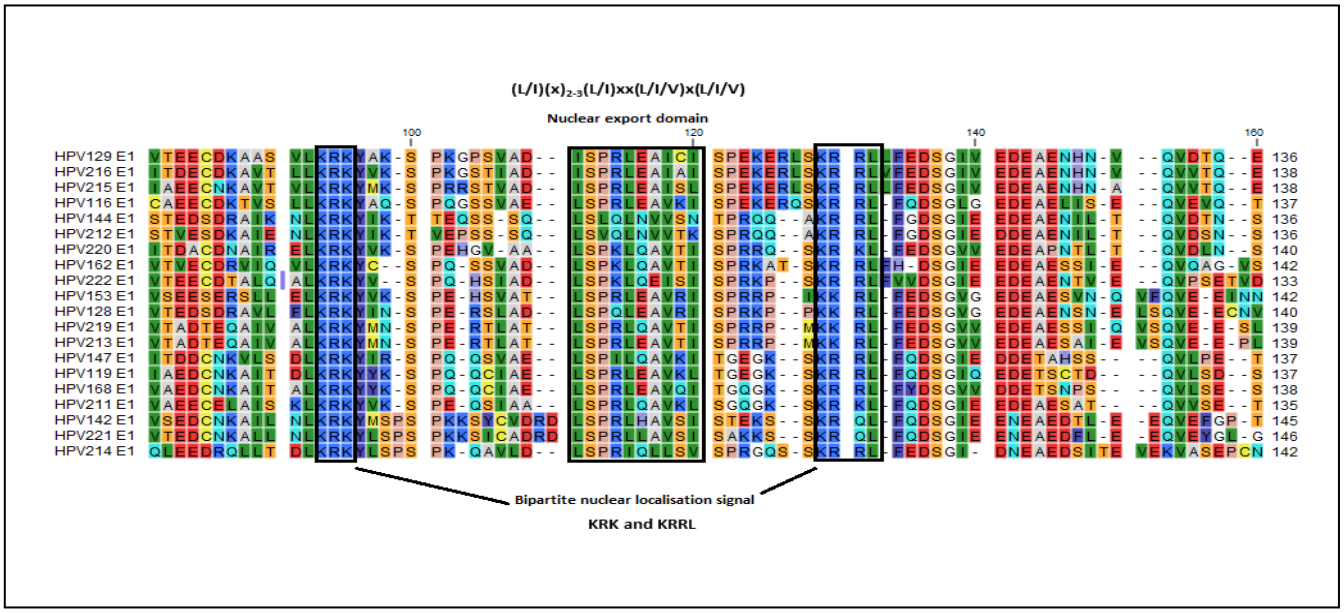
Supplementary Figure 2.12 Alignment of the E7 proteins of the novel HPVs and closely related HPVs.

The positions of the pRB binding domain and Zinc finger binding domains in E7 are indicated by the black boxes.



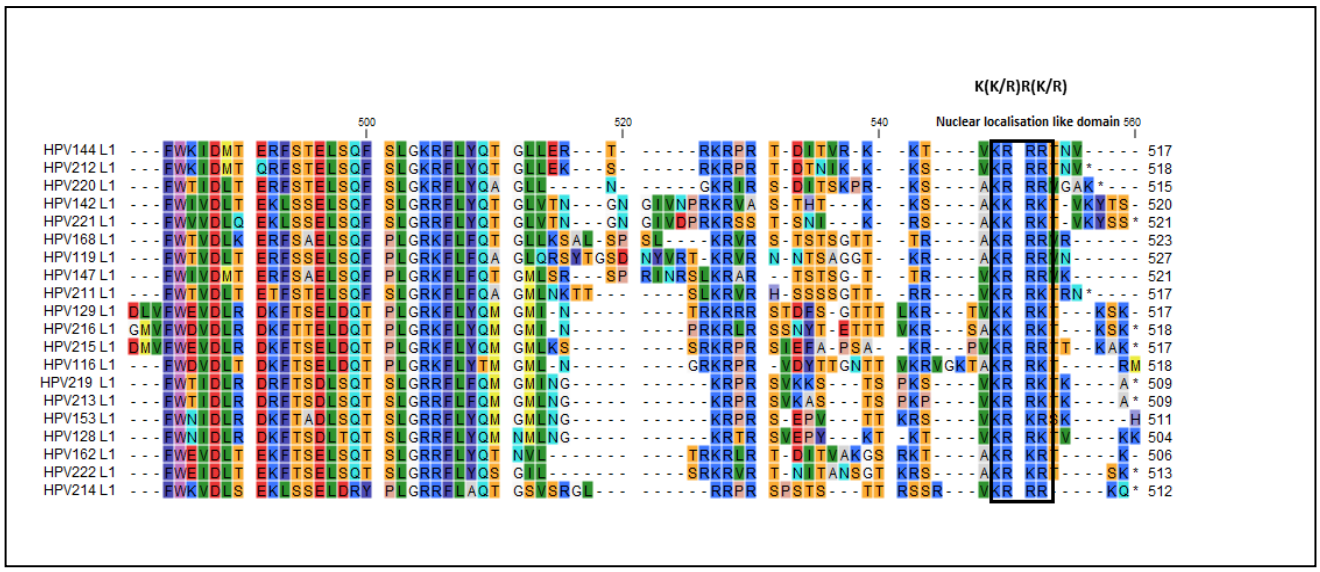
Supplementary Figure 2.13 Alignment of the E1 proteins of the novel HPVs and closely related HPVs.

The positions of the ATP binding sites are indicated by the black boxes



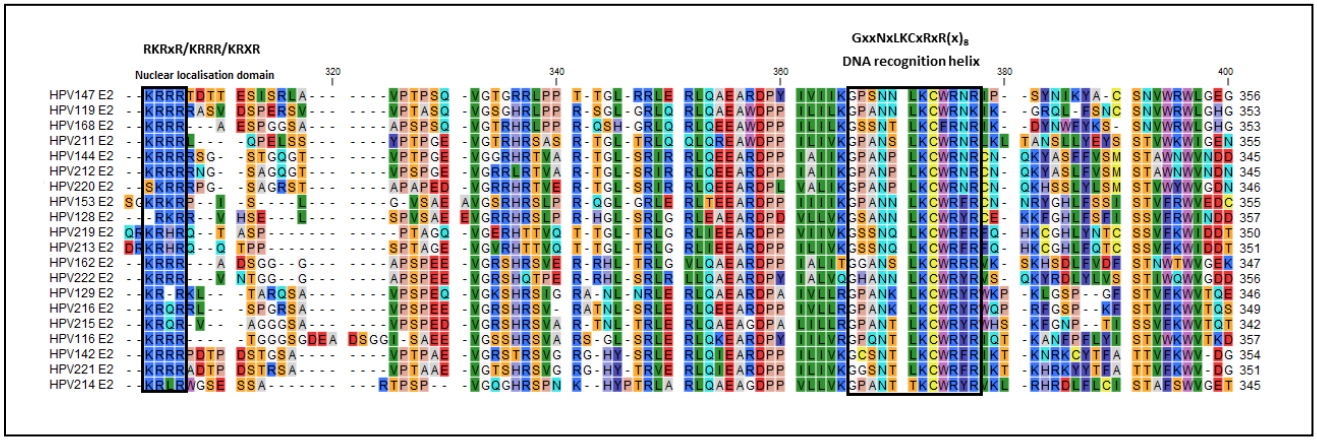
Supplementary Figure 2.14 Alignment of the E1 proteins of the novel HPVs and closely related HPVs.

The positions of the bipartite NLS and NES sites are indicated by the black boxes.



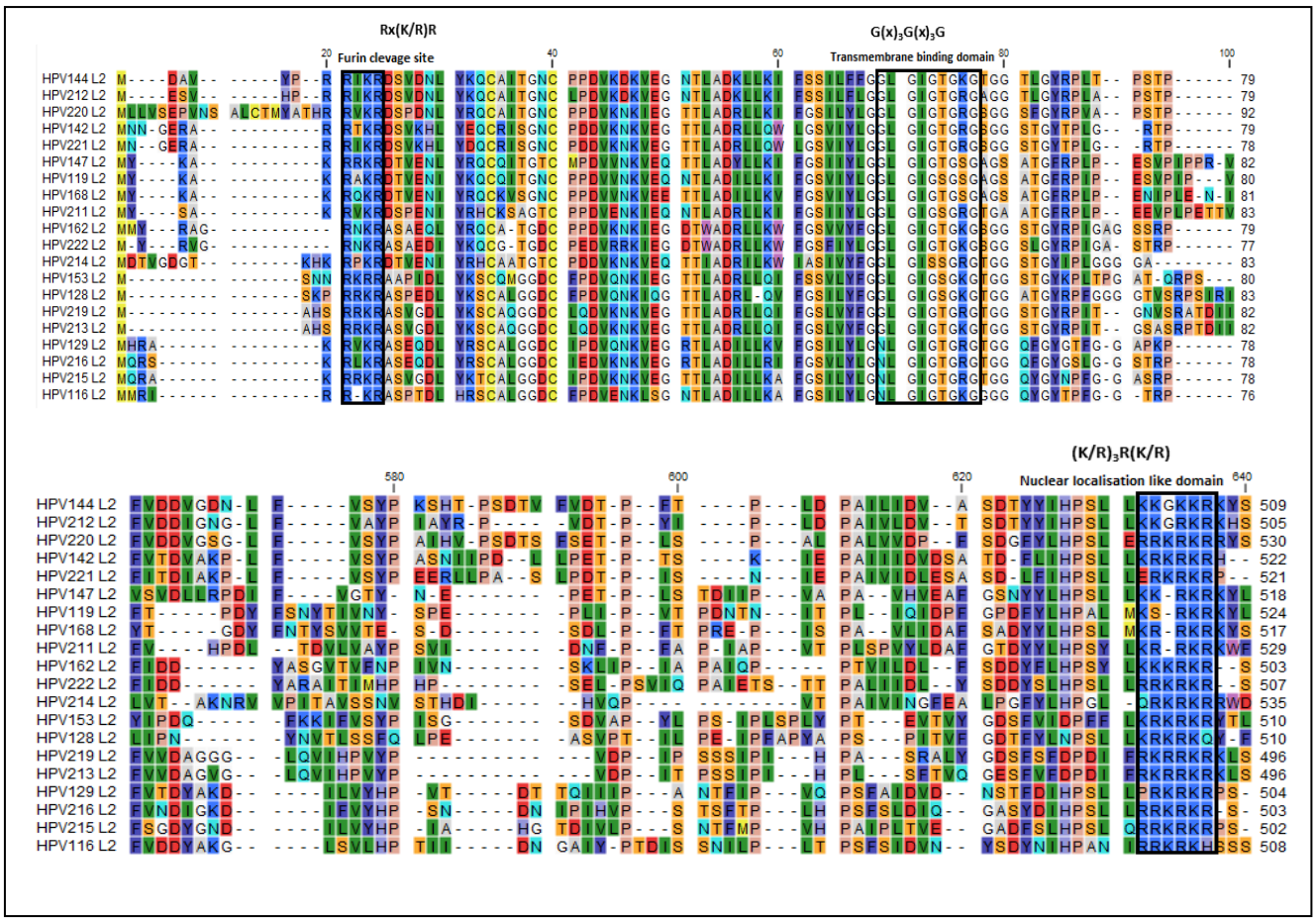
Supplementary Figure 2.15 Alignment of the L1 proteins of the novel HPVs and closely related HPVs.

The positions of the NLS sites are indicated by the black boxes.



Supplementary Figure 2.16 Alignment of the E2 proteins of the novel HPVs and closely related HPVs.

The positions of the NLS sites and DNA recognition helix are indicated by the black boxes.



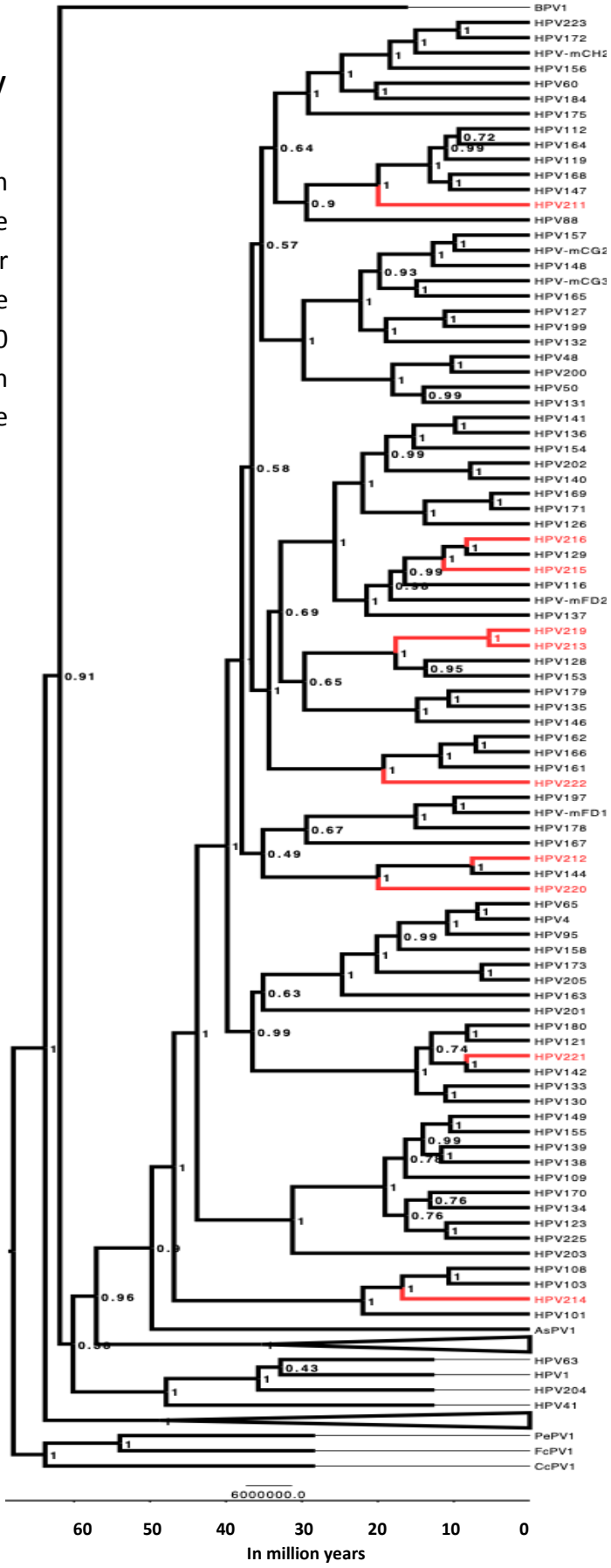
Supplementary Figure 2.17 Alignment of the L2 proteins of the novel HPVs and closely related HPVs.

The positions of the NLS sites, transmembrane binding site and the Furin cleavage sites are indicated by the black boxes.

Appendix 2: Appendix to Chapter 3

Supplementary Figure 3. 5 Posterior support values of PV divergence estimates.

Classification was based on (Bernard et al., 2010, de Villiers et al., 2004). Posterior support values are shown, the nodes corresponding to the 10 HPV types are presented in Table 3.4, the novel types are indicated in red.



Supplementary Figure 3. 6A- F (below) Six Tables showing the Shimodaira-Hasegawa test results using W-IQ-Tree.

Tree Key

1-E1 tree 2-E2 tree 3-E4 tree 4-E7 tree 5- L1 tree 6-L2 tree

```
USER TREES
----------

See E1.clw.trees for trees with branch lengths.

WARNING: Too few replicates for AU test. At least -zb 10000 for reliable results!

Tree      logL    deltaL  bp-RELL    p-KH     p-SH    p-WKH    p-WSH       c-ELW       p-AU
--------------------------------------------------------------------------------------------
  1 -96195.23188       0       1 +      1 +      1 +      1 +      1 +         1 +    0.999 +
  2 -96377.39456  182.16       0 -      0 -  0.008 -      0 -      0 -  7.34e-33 - 0.000747 -
  3 -96834.37295  639.14       0 -      0 -      0 -      0 -      0 - 3.82e-203 - 0.000845 -
  4 -97243.96632  1048.7       0 -      0 -      0 -      0 -      0 -         0 - 1.77e-58 -
  5 -96576.44197  381.21       0 -      0 -      0 -      0 -      0 - 3.76e-104 - 4.21e-05 -
  6 -96414.70358  219.47       0 -      0 -  0.002 -      0 -      0 -  1.23e-40 - 2.36e-09 -

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
p-WKH   : p-value of weighted KH test.
p-WSH   : p-value of weighted SH test.
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.
```

A

```
See E2.clw.trees for trees with branch lengths.

WARNING: Too few replicates for AU test. At least -zb 10000 for reliable results!

Tree      logL    deltaL  bp-RELL    p-KH     p-SH    p-WKH    p-WSH       c-ELW       p-AU
--------------------------------------------------------------------------------------------
  1 -73154.37589  137.43       0 -      0 -  0.023 -      0 -      0 -  1.34e-27 - 8.32e-05 -
  2  -73016.9446       0       1 +      1 +      1 +      1 +      1 +         1 +        1 +
  3 -73305.29284  288.35       0 -      0 -      0 -      0 -      0 -  1.26e-75 - 2.84e-07 -
  4 -73921.86909  904.92       0 -      0 -      0 -      0 -      0 - 3.41e-299 - 6.72e-97 -
  5 -73290.15172  273.21       0 -      0 -      0 -      0 -      0 -  5.91e-70 - 5.74e-44 -
  6 -73241.89999  224.96       0 -      0 -      0 -      0 -      0 -  1.77e-60 - 5.05e-07 -

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
p-WKH   : p-value of weighted KH test.
p-WSH   : p-value of weighted SH test.
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.

TIME STAMP
----------

Date and time: Mon Jan 28 13:06:24 2019
Total CPU time used: 10.6 seconds (0h:0m:10s)
Total wall-clock time used: 11.8 seconds (0h:0m:11s)
```

B

```
See E4.clw.trees for trees with branch lengths.

WARNING: Too few replicates for AU test. At least -zb 10000 for reliable results!

Tree      logL    deltaL  bp-RELL    p-KH     p-SH    p-WKH    p-WSH       c-ELW       p-AU
--------------------------------------------------------------------------------------------
  1 -29182.09554  79.743       0 -      0 -  0.012 -      0 -  0.004 -  0.000433 -   0.0217 -
  2 -29167.34655  64.994   0.002 -  0.003 -   0.02 -  0.003 -  0.016 -   0.00211 -   0.0104 -
  3 -29102.35229       0   0.998 +  0.997 +      1 +  0.997 +      1 +     0.997 +    0.992 +
  4 -29414.21427  311.86       0 -      0 -      0 -      0 -      0 -  8.65e-85 - 5.78e-114 -
  5 -29232.50125  130.15       0 -      0 -      0 -      0 -      0 -  1.15e-23 - 6.36e-35 -
  6 -29187.17923  84.827       0 -      0 -  0.004 -      0 -      0 -  1.09e-10 -  0.00483 -

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
p-WKH   : p-value of weighted KH test.
p-WSH   : p-value of weighted SH test.
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.

TIME STAMP
----------

Date and time: Mon Jan 28 13:10:19 2019
Total CPU time used: 7.42 seconds (0h:0m:7s)
Total wall-clock time used: 8.03 seconds (0h:0m:8s)
```

C

```
See E7.clw.trees for trees with branch lengths.

WARNING: Too few replicates for AU test. At least -zb 10000 for reliable results!

Tree      logL    deltaL  bp-RELL    p-KH     p-SH    p-WKH    p-WSH       c-ELW       p-AU
--------------------------------------------------------------------------------------------
  1  -17499.0381  59.737   0.004 -  0.016 -  0.017 -  0.016 -  0.064 +   0.00429 -   0.0117 -
  2 -17510.31705 71.016   0.001 -  0.007 -  0.007 -  0.007 -  0.031 -  0.000919 -  0.00182 -
  3 -17577.72981 138.43       0 -      0 -      0 -      0 -      0 -  1.78e-20 -    0.014 -
  4 -17439.30074      0   0.946 +  0.952 +      1 +  0.952 +  0.991 +     0.944 +    0.945 +
  5 -17478.72361 39.423   0.041 +  0.048 -  0.119 +  0.048 -  0.189 +     0.043 +    0.118 +
  6 -17499.15756 59.857   0.008 -  0.022 -  0.025 -  0.022 -  0.063 +   0.00759 -   0.0321 -

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
p-WKH   : p-value of weighted KH test.
p-WSH   : p-value of weighted SH test.
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.

TIME STAMP
----------

Date and time: Mon Jan 28 13:14:09 2019
Total CPU time used: 2.08 seconds (0h:0m:2s)
Total wall-clock time used: 2.51 seconds (0h:0m:2s)
```

D

```
See L1.clw.trees for trees with branch lengths.

WARNING: Too few replicates for AU test. At least -zb 10000 for reliable results!

Tree      logL    deltaL  bp-RELL    p-KH     p-SH    p-WKH    p-WSH       c-ELW       p-AU
--------------------------------------------------------------------------------------------
  1  -86530.3421  188.69       0 -      0 -      0 -      0 -      0 -  1.32e-25 - 7.65e-05 -
  2 -86560.16506 218.51       0 -      0 -      0 -      0 -      0 -   4.1e-40 - 6.59e-48 -
  3 -86814.86187 473.21       0 -      0 -      0 -      0 -      0 - 1.61e-133 - 0.000267 -
  4 -87241.31128 899.66       0 -      0 -      0 -      0 -      0 - 3.53e-301 - 5.98e-65 -
  5 -86341.65032      0       1 +      1 +      1 +      1 +      1 +         1 +        1 +
  6 -86491.47873 149.83       0 -      0 -  0.005 -      0 -      0 -  9.88e-12 - 1.23e-10 -

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
p-WKH   : p-value of weighted KH test.
p-WSH   : p-value of weighted SH test.
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.

TIME STAMP
----------

Date and time: Mon Jan 28 13:16:10 2019
Total CPU time used: 14.4 seconds (0h:0m:14s)
Total wall-clock time used: 15.6 seconds (0h:0m:15s)
```

E

```
See L2.clw.trees for trees with branch lengths.

WARNING: Too few replicates for AU test. At least -zb 10000 for reliable results!

Tree      logL    deltaL  bp-RELL    p-KH     p-SH    p-WKH    p-WSH       c-ELW       p-AU
--------------------------------------------------------------------------------------------
  1 -104837.6379  226.34       0 -      0 -      0 -      0 -      0 -  2.42e-51 -  1.3e-52 -
  2 -104831.5886  220.29       0 -      0 -      0 -      0 -      0 -  1.84e-50 - 8.38e-12 -
  3 -105090.0747  478.78       0 -      0 -      0 -      0 -      0 - 1.49e-149 - 4.23e-06 -
  4 -105698.3635  1087.1       0 -      0 -      0 -      0 -      0 -         0 -  0.00124 -
  5 -104841.0161  229.72       0 -      0 -      0 -      0 -      0 -  1.32e-42 - 2.01e-13 -
  6 -104611.2977       0       1 +      1 +      1 +      1 +      1 +         1 +        1 +

deltaL  : logL difference from the maximal logl in the set.
bp-RELL : bootstrap proportion using RELL method (Kishino et al. 1990).
p-KH    : p-value of one sided Kishino-Hasegawa test (1989).
p-SH    : p-value of Shimodaira-Hasegawa test (2000).
p-WKH   : p-value of weighted KH test.
p-WSH   : p-value of weighted SH test.
c-ELW   : Expected Likelihood Weight (Strimmer & Rambaut 2002).
p-AU    : p-value of approximately unbiased (AU) test (Shimodaira, 2002).

Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.
All tests performed 1000 resamplings using the RELL method.

TIME STAMP
----------

Date and time: Mon Jan 28 13:18:29 2019
Total CPU time used: 21.1 seconds (0h:0m:21s)
Total wall-clock time used: 22.7 seconds (0h:0m:22s)
```

F

References

2007., I. M. 2007. Human Papillomavirus. *IARC Monographs on the Evaluation of Carcinogenic Risk s to Humans,* 90**,** 61-62.

AGALLIU, I., GAPSTUR, S., CHEN, Z., WANG, T., ANDERSON, R. L., TERAS, L., KREIMER, A. R., HAYES, R. B., FREEDMAN, N. D. & BURK, R. D. 2016. Associations of Oral alpha-, beta-, and gamma-Human Papillomavirus Types With Risk of Incident Head and Neck Cancer. *JAMA Oncol*.

AKOGBE, G. O., AJIDAHUN, A., SIRAK, B., ANIC, G. M., PAPENFUSS, M. R., FULP, W. J., LIN, H. Y., ABRAHAMSEN, M., VILLA, L. L., LAZCANO-PONCE, E., QUITERIO, M., SMITH, D., SCHABATH, M. B., SALMERON, J. & GIULIANO, A. R. 2012. Race and prevalence of human papillomavirus infection among men residing in Brazil, Mexico and the United States. *Int J Cancer,* 131**,** E282-91.

AKSOY, P., GOTTSCHALK, E. Y. & MENESES, P. I. 2017. HPV entry into cells. *Mutat Res Rev Mutat Res,* 772**,** 13-22.

AMEUR, A., MEIRING, T. L., BUNIKIS, I., HAGGQVIST, S., LINDAU, C., LINDBERG, J. H., GUSTAVSSON, I., MBULAWA, Z. Z., WILLIAMSON, A. L. & GYLLENSTEN, U. 2014a. Comprehensive profiling of the vaginal microbiome in HIV positive women using massive parallel semiconductor sequencing. *Scientific reports,* 4**,** 4398.

AMEUR, A., MEIRING, T. L., BUNIKIS, I., HAGGQVIST, S., LINDAU, C., LINDBERG, J. H., GUSTAVSSON, I., MBULAWA, Z. Z., WILLIAMSON, A. L. & GYLLENSTEN, U. 2014b. Comprehensive profiling of the vaginal microbiome in HIV positive women using massive parallel semiconductor sequencing. *Sci Rep***,** 4:4398.

ANDREWS, P. 1992. Evolution and environment in the Hominoidea. *Nature,* 360**,** 641-6.

ANDROPHY, E. J., LOWY, D. R. & SCHILLER, J. T. 1987. Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature,* 325**,** 70-3.

ANGULO, M. & CARVAJAL-RODRIGUEZ, A. 2007. Evidence of recombination within human alpha-papillomavirus. *Virol J,* 4**,** 33.

ANISIMOVA, M. & GASCUEL, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology,* 55**,** 539-552.

ANTONSSON, A., FORSLUND, O., EKBERG, H., STERNER, G. & HANSSON, B. G. 2000. The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. *Journal of virology,* 74**,** 11636-41.

ARBYN, M., TOMMASINO, M., DEPUYDT, C. & DILLNER, J. 2014. Are twenty human papillomavirus types causing cervical cancer? *J Pathol*.

ARROYO MUHR, L. S., HULTIN, E., BZHALAVA, D., EKLUND, C., LAGHEDEN, C., EKSTROM, J., JOHANSSON, H., FORSLUND, O. & DILLNER, J. 2014. Human papillomavirus type 197 is commonly present in skin tumors. *Int J Cancer*.

ASHRAFI, G. H., HAGHSHENAS, M., MARCHETTI, B. & CAMPO, M. S. 2006. E5 protein of human papillomavirus 16 downregulates HLA class I and interacts with the heavy chain via its first hydrophobic domain. *Int J Cancer,* 119**,** 2105-12.

BEDARD, K. M., UNDERBRINK, M. P., HOWIE, H. L. & GALLOWAY, D. A. 2008. The E6 oncoproteins from human betapapillomaviruses differentially activate telomerase through an E6AP-dependent mechanism and prolong the lifespan of primary keratinocytes. *J Virol,* 82**,** 3894-902.

BEDROSIAN, C. L. & BASTIA, D. 1990. The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the nonconserved core sequence in binding site affinity. *Virology,* 174**,** 557-75.

BERGVALL, M., MELENDY, T. & ARCHAMBAULT, J. 2013a. The E1 proteins. *Virology,* 445**,** 35-56.

BERGVALL, M., MELENDY, T. & ARCHAMBAULT, J. 2013b. The E1 proteins. *Virology,* 445**,** 35-56.

BERNARD, H.-U. 2013a. Regulatory elements in the viral genome. *Virology,* 445**,** 197-204.

BERNARD, H. U. 2013b. Regulatory elements in the viral genome. *Virology,* 445**,** 197-204.

BERNARD, H. U., BURK, R. D., CHEN, Z., VAN DOORSLAER, K., ZUR HAUSEN, H. & DE VILLIERS, E. M. 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 401(1)**,** 70-79.

BERNARD, H. U., CALLEJA-MACIAS, I. E. & DUNN, S. T. 2006. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer,* 118**,** 1071-6.

BERNHOFER, M., GOLDBERG, T., WOLF, S., AHMED, M., ZAUGG, J., BODEN, M. & ROST, B. 2018. NLSdb—major update for database of nuclear localization signals and nuclear export signals. *Nucleic Acids Research,* 46**,** D503-D508.

BISHOP, B., DASGUPTA, J. & CHEN, X. S. 2007. Structure-based engineering of papillomavirus major capsid l1: controlling particle assembly. *Virol J,* 4**,** 3.

BOLATTI, E. M., CHOUHY, D., CASAL, P. E., PEREZ, G. R., STELLA, E. J., SANCHEZ, A., GOROSITO, M., BUSSY, R. F. & GIRI, A. A. 2016. Characterization of novel human papillomavirus types 157, 158 and 205 from healthy skin and recombination analysis in genus gamma-Papillomavirus. *Infect Genet Evol,* 42**,** 20-29.

BOLATTI, E. M., HOSNJAK, L., CHOUHY, D., RE-LOUHAU, M. F., CASAL, P. E., BOTTAI, H., KOCJAN, B. J., STELLA, E. J., GOROSITO, M. D., SANCHEZ, A., BUSSY, R. F., POLJAK, M. & GIRI, A. A. 2018. High prevalence of Gammapapillomaviruses (Gamma-PVs) in pre-malignant cutaneous lesions of immunocompetent individuals using a new broad-spectrum primer system, and identification of HPV210, a novel Gamma-PV type. *Virology,* 525**,** 182-191.

BORDEAUX, J., FORTE, S., HARDING, E., DARSHAN, M. S., KLUCEVSEK, K. & MOROIANU, J. 2006. The l2 minor capsid protein of low-risk human papillomavirus type 11 interacts with host nuclear import receptors and viral DNA. *J Virol,* 80**,** 8259-62.

BOTTALICO, D., CHEN, Z., DUNNE, A., OSTOLOZA, J., MCKINNEY, S., SUN, C., SCHLECHT, N. F., FATAHZADEH, M., HERRERO, R., SCHIFFMAN, M. & BURK, R. D. 2011a. The oral cavity contains abundant known and novel human papillomaviruses from the Betapapillomavirus and Gammapapillomavirus genera. *J Infect Dis,* 204**,** 787-92.

BOTTALICO, D., CHEN, Z., DUNNE, A., OSTOLOZA, J., MCKINNEY, S., SUN, C., SCHLECHT, N. F., FATAHZADEH, M., HERRERO, R., SCHIFFMAN, M. & BURK, R. D. 2011b. The oral cavity contains abundant known and novel human papillomaviruses from the Betapapillomavirus and Gammapapillomavirus genera. *The Journal of infectious diseases,* 204**,** 787-92.

BOTTALICO, D., CHEN, Z., DUNNE, A., OSTOLOZA, J., MCKINNEY, S., SUN, C., SCHLECHT, N. F., FATAHZADEH, M., HERRERO, R., SCHIFFMAN, M. & BURK, R. D. 2011c. The Oral Cavity Contains Abundant Known and Novel Human Papillomaviruses From the Betapapillomavirus and Gammapapillomavirus Genera. *The Journal of Infectious Diseases,* 204**,** 787-792.

BOTTALICO, D., CHEN, Z., KOCJAN, B. J., SEME, K., POLJAK, M. & BURK, R. D. 2012. Characterization of human papillomavirus type 120: a novel betapapillomavirus with tropism for multiple anatomical niches. *J Gen Virol,* 93**,** 1774-9.

BOUWES BAVINCK, J. N., NEALE, R. E., ABENI, D., EUVRARD, S., GREEN, A. C., HARWOOD, C. A., DE KONING, M. N., NALDI, L., NINDL, I., PAWLITA, M., PFISTER, H., PROBY, C. M., QUINT, W. G., TER SCHEGGET, J., WATERBOER, T., WEISSENBORN, S. & FELTKAMP, M. C. 2010. Multicenter study of the association between betapapillomavirus infection and cutaneous squamous cell carcinoma. *Cancer Res,* 2010 Dec 1;70(23):9777-86. doi**,** 10.1158/0008-5472.CAN-10-0352.

BRANCACCIO, R. N., ROBITAILLE, A., DUTTA, S., ROLLISON, D. E., FISCHER, N., GRUNDHOFF, A., TOMMASINO, M. & GHEIT, T. 2017. Complete Genome Sequence of a Novel Human Gammapapillomavirus Isolated from Skin. *Genome Announcements,* 5.

BRAVO, I. G. & ALONSO, A. 2004a. Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. *J Virol,* 78**,** 13613-26.

BRAVO, I. G. & ALONSO, Á. 2004b. Mucosal Human Papillomaviruses Encode Four Different E5 Proteins Whose Chemistry and Phylogeny Correlate with Malignant or Benign Growth. *Journal of Virology,* 78**,** 13613-13626.

BRAVO, I. G. & FELEZ-SANCHEZ, M. 2015. Papillomaviruses: Viral evolution, cancer and evolutionary medicine. *Evol Med Public Health,* 2015**,** 32-51.

BRONNIMANN, M. P., CHAPMAN, J. A., PARK, C. K. & CAMPOS, S. K. 2013. A transmembrane domain and GxxxG motifs within L2 are essential for papillomavirus infection. *J Virol,* 87**,** 464-73.

BURK, R. D., HARARI, A. & CHEN, Z. 2013. Human papillomavirus genome variants. *Virology* 445**,** 232-43.

BZHALAVA, D., EKLUND, C. & DILLNER, J. 2015. International standardization and classification of human papillomavirus types. *Virology***,** 2015 Jan 8;476C:341-344.

BZHALAVA, D., GUAN, P., FRANCESCHI, S., DILLNER, J. & CLIFFORD, G. 2013. A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types. *Virology,* 445**,** 224-31.

CALLEJA-MACIAS, I. E., KALANTARI, M., ALLAN, B., WILLIAMSON, A. L., CHUNG, L. P., COLLINS, R. J., ZUNA, R. E., DUNN, S. T., ORTIZ-LOPEZ, R., BARRERA-SALDANA, H. A., CUBIE, H. A., CUSCHIERI, K., VILLA, L. L. & BERNARD, H. U. 2005. Papillomavirus subtypes are natural and old taxa: phylogeny of human papillomavirus types 44 and 55 and 68a and -b. *J Virol,* 79**,** 6565-9.

CARRILLO-GARCIA, A., PONCE-DE-LEON-ROSALES, S., CANTU-DE-LEON, D., FRAGOSO-ONTIVEROS, V., MARTINEZ-RAMIREZ, I., OROZCO-COLIN, A., MOHAR, A. & LIZANO, M. 2014. Impact of human papillomavirus coinfections on the risk of high-grade squamous intraepithelial lesion and cervical cancer. *Gynecol Oncol,* 134**,** 534-9.

CARTER, J. J., WIPF, G. C., MADELEINE, M. M., SCHWARTZ, S. M., KOUTSKY, L. A. & GALLOWAY, D. A. 2006. Identification of human papillomavirus type 16 L1 surface loops required for neutralization by human sera. *J Virol,* 80**,** 4664-72.

CARVAJAL-RODRIGUEZ, A. 2008. Detecting recombination and diversifying selection in human alpha-papillomavirus. *Infect Genet Evol,* 8**,** 689-92.

CASTOE, T. A., DE KONING, A. P., KIM, H. M., GU, W., NOONAN, B. P., NAYLOR, G., JIANG, Z. J., PARKINSON, C. L. & POLLOCK, D. D. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci U S A,* 106**,** 8986-91.

CERQUEIRA, C. & SCHILLER, J. T. 2017. Papillomavirus assembly: an overview and perspectives. *Virus research,* 231**,** 103-107.

CHEN, S. L. & MOUNTS, P. 1990. Transforming activity of E5a protein of human papillomavirus type 6 in NIH 3T3 and C127 cells. *Journal of virology,* 64**,** 3226-3233.

CHEN, X. S., GARCEA, R. L., GOLDBERG, I., CASINI, G. & HARRISON, S. C. 2000. Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. *Mol Cell,* 5**,** 557-67.

CHEN, Z., DE FREITAS, L. B. & BURK, R. D. 2015. Evolution and classification of oncogenic human papillomavirus types and variants associated with cervical cancer. *Methods Mol Biol,* 1249**,** 3-26.

CHEN, Z., DESALLE, R., SCHIFFMAN, M., HERRERO, R. & BURK, R. D. 2009. Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. *J Virol,* 83**,** 1443-55.

CHEN, Z., SCHIFFMAN, M., HERRERO, R. & BURK, R. D. 2007a. Identification and characterization of two novel human papillomaviruses (HPVs) by overlapping PCR: HPV102 and HPV106. *J Gen Virol,* 88**,** 2952-5.

CHEN, Z., SCHIFFMAN, M., HERRERO, R., DESALLE, R. & BURK, R. D. 2007b. Human papillomavirus (HPV) types 101 and 103 isolated from cervicovaginal cells lack an E6 open reading frame (ORF) and are related to gamma-papillomaviruses. *Virol J,* 360**,** 447-53.

CHEN, Z., TERAI, M., FU, L., HERRERO, R., DESALLE, R. & BURK, R. D. 2005. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. *J Virol,* 79**,** 7014-23.

COLON-LOPEZ, V., SHIELS, M. S., MACHIN, M., ORTIZ, A. P., STRICKLER, H., CASTLE, P. E., PFEIFFER, R. M. & ENGELS, E. A. 2018. Anal Cancer Risk Among People With HIV Infection in the United States. *J Clin Oncol,* 36**,** 68-75.

CONWAY, M. J. & MEYERS, C. 2009. Replication and assembly of human papillomaviruses. *J Dent Res***,** 2009 Apr;88(4):307-17.

CORNALL, A. M., ROBERTS, J. M., GARLAND, S. M., HILLMAN, R. J., GRULICH, A. E. & TABRIZI, S. N. 2013. Anal and perianal squamous carcinomas and high-grade intraepithelial lesions exclusively associated with "low-risk" HPV genotypes 6 and 11. *International Journal of Cancer,* 133**,** 2253-2258.

DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods,* 9**,** 772.

DAY, P. M., LOWY, D. R. & SCHILLER, J. T. 2008. Heparan sulfate-independent cell binding and infection with furin-precleaved papillomavirus capsids. *J Virol,* 82**,** 12565-8.

DE MARTEL, C. 2017. Worldwide burden of cancer attributable to HPV by site, country and HPV type. 141**,** 664-70.

DE OLIVEIRA, C. M., BRAVO, I. G., SANTIAGO E SOUZA, N. C., GENTA, M. L., FREGNANI, J. H., TACLA, M., CARVALHO, J. P., LONGATTO-FILHO, A. & LEVI, J. E. 2015. High-level of viral genomic diversity in cervical cancers: A Brazilian study on human papillomavirus type 16. *Infect Genet Evol,* 34**,** 44-51.

DE VILLIERS, E. M. 2013. Cross-roads in the classification of papillomaviruses. *Virology,* 445**,** 2-10.

DE VILLIERS, E. M., FAUQUET, C., BROKER, T. R., BERNARD, H. U. & ZUR HAUSEN, H. 2004. Classification of papillomaviruses. *Virology,* 324**,** 17-27.

DE VILLIERS, E. M. & GUNST, K. 2009. Characterization of seven novel human papillomavirus types isolated from cutaneous tissue, but also present in mucosal lesions. *J Gen Virol,* 90**,** 1999-2004.

DEFILIPPIS, V. R., AYALA, F. J. & VILLARREAL, L. P. 2002. Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes. *J Mol Evol,* 55**,** 491-9.

DEL PINO, M., BLEEKER, M. C., QUINT, W. G., SNIJDERS, P. J., MEIJER, C. J. & STEENBERGEN, R. D. 2012. Comprehensive analysis of human papillomavirus prevalence and the potential role of low-risk types in verrucous carcinoma. *Mod Pathol,* 25**,** 1354-63.

DELURY, C. P., MARSH, E., JAMES, C. D., BOON, S. S., BANKS, L., KNIGHT, G. L. & ROBERTS, S. 2013. The role of protein kinase A regulation of the E6 PDZ-binding domain during the differentiation-dependent life cycle of human papillomavirus type 18. *Journal of Virology*.

DENNY, L., ADEWOLE, I., ANORLU, R., DREYER, G., MOODLEY, M., SMITH, T., SNYMAN, L., WIREDU, E., MOLIJN, A., QUINT, W., RAMAKRISHNAN, G. & SCHMIDT, J. 2014. Human papillomavirus prevalence and type distribution in invasive cervical cancer in sub-Saharan Africa. *Int J Cancer,* 134**,** 1389-98.

DI BONITO, P., DELLA LIBERA, S., PETRICCA, S., IACONELLI, M., SANGUINETTI, M., GRAFFEO, R., ACCARDI, L. & LA ROSA, G. 2015. A large spectrum of alpha and beta papillomaviruses are detected in human stool samples. *J Gen Virol,* 96**,** 607-13.

DIGIUSEPPE, S., BIENKOWSKA-HABA, M., GUION, L. G. & SAPP, M. 2017. Cruising the cellular highways: How human papillomavirus travels from the surface to the nucleus. *Virus Research,* 231**,** 1-9.

DIMAIO, D. & PETTI, L. M. 2013. The E5 proteins. *Virology,* 445**,** 99-114.

DOORBAR, J. 2005. The papillomavirus life cycle. *J Clin Virol,* 32 Suppl 1**,** S7-15.

DOORBAR, J. 2013. The E4 protein; structure, function and patterns of expression. *Virology,* 445**,** 80-98.

DOORBAR, J., EGAWA, N., GRIFFIN, H., KRANJEC, C. & MURAKAMI, I. 2015. Human papillomavirus molecular biology and disease association. *Rev Med Virol,* 25 Suppl 1**,** 2-23.

DOORBAR, J., ELY, S., STERLING, J., MCLEAN, C. & CRAWFORD, L. 1991. Specific interaction between HPV-16 E1–E4 and cytokeratins results in collapse of the epithelial cell intermediate filament network. *Nature,* 352**,** 824.

DOORBAR, J., MEDCALF, E. & NAPTHINE, S. 1996. Analysis of HPV1 E4 Complexes and Their Association with Keratinsin Vivo. *Virology,* 218**,** 114-126.

DOORBAR, J., QUINT, W., BANKS, L., BRAVO, I. G., STOLER, M., BROKER, T. R. & STANLEY, M. A. 2012. The biology and life-cycle of human papillomaviruses. *Vaccine,* 30 Suppl 5**,** F55-70.

DRUMMOND, A. J., SUCHARD, M. A., XIE, D. & RAMBAUT, A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution,* 29**,** 1969-1973.

DUBE MANDISHORA, R. S., GJOTTERUD, K. S., LAGSTROM, S., STRAY-PEDERSEN, B., DURI, K., CHIN'OMBE, N., NYGARD, M., CHRISTIANSEN, I. K., AMBUR, O. H., CHIRENJE, M. Z. & ROUNGE, T. B. 2018. Intra-host sequence variability in human papillomavirus. *Papillomavirus Res,* 5**,** 180-191.

DUNNE, E. F., NIELSON, C. M., STONE, K. M., MARKOWITZ, L. E. & GIULIANO, A. R. 2006. Prevalence of HPV infection among men: A systematic review of the literature. *J Infect Dis,* 194**,** 1044-57.

DUTTA, S., ROBITAILLE, A., ROLLISON, D. E., TOMMASINO, M. & GHEIT, T. 2017. Complete Genome Sequence of a Novel Human Betapapillomavirus Isolated from a Skin Sample. *Genome Announcements,* 5.

DZIDUSZKO, A. & OZBUN, M. A. 2013. Annexin A2 and S100A10 regulate human papillomavirus type 16 entry and intracellular trafficking in human keratinocytes. *J Virol,* 87**,** 7502-15.

EDGAR, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5**,** 113.

EDGAR, R. C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

EGAWA, K. 1994. New types of human papillomaviruses and intracytoplasmic inclusion bodies: a classification of inclusion warts according to clinical features, histology and associated HPV types. *British Journal of Dermatology,* 130**,** 158-166.

EGAWA, N. & DOORBAR, J. 2017. The low-risk papillomaviruses. *Virus Res,* 231**,** 119-127.

EGAWA, N., EGAWA, K., GRIFFIN, H. & DOORBAR, J. 2015. Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses,* 7**,** 3863-90.

EKSTRÖM, J., BZHALAVA, D., SVENBACK, D., FORSLUND, O. & DILLNER, J. 2011. High throughput sequencing reveals diversity of Human Papillomaviruses in cutaneous lesions. *Int J Cancer,* 129**,** 2643-2650.

EKSTROM, J., MUHR, L. S., BZHALAVA, D., SODERLUND-STRAND, A., HULTIN, E., NORDIN, P., STENQUIST, B., PAOLI, J., FORSLUND, O. & DILLNER, J. 2013a. Diversity of human papillomaviruses in skin lesions. *Virology,* 447**,** 300-11.

EKSTROM, J., MUHR, L. S., BZHALAVA, D., SODERLUND-STRAND, A., HULTIN, E., NORDIN, P., STENQUIST, B., PAOLI, J., FORSLUND, O. & DILLNER, J. 2013b. Diversity of human papillomaviruses in skin lesions. *Virology,* 2013 Dec;447(1-2):300-11. doi**,** 10.1016/j.virol.2013.09.010.

EVANDER, M., FRAZER, I. H., PAYNE, E., QI, Y. M., HENGST, K. & MCMILLAN, N. A. 1997. Identification of the alpha6 integrin as a candidate receptor for papillomaviruses. *J Virol,* 71**,** 2449-56.

FISCHER, U., HUBER, J., BOELENS, W. C., MATTAJ, I. W. & LUHRMANN, R. 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell,* 82**,** 475-83.

FONG, J. H. & MARCHLER-BAUER, A. 2008. Protein subfamily assignment using the Conserved Domain Database. *BMC Research Notes,* 1**,** 114.

FORMAN, D., DE MARTEL, C., LACEY, C. J., SOERJOMATARAM, I., LORTET-TIEULENT, J., BRUNI, L., VIGNAT, J., FERLAY, J., BRAY, F., PLUMMER, M. & FRANCESCHI, S. 2012. Global burden of human papillomavirus and related diseases. *Vaccine,* 30 Suppl 5**,** F12-23.

FORSLUND, O., ANTONSSON, A., NORDIN, P., STENQUIST, B. & HANSSON, B. G. 1999. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J Gen Virol,* 80 (Pt 9)**,** 2437-43.

FORSLUND, O., JOHANSSON, H., MADSEN, K. G. & KOFOED, K. 2013a. The nasal mucosa contains a large spectrum of human papillomavirus types from the Betapapillomavirus and Gammapapillomavirus genera. *J Infect Dis,* 2013 Oct 15;208(8):1335-41. doi**,** 10.1093/infdis/jit326.

FORSLUND, O., JOHANSSON, H., MADSEN, K. G. & KOFOED, K. 2013b. The nasal mucosa contains a large spectrum of human papillomavirus types from the Betapapillomavirus and Gammapapillomavirus genera. *The Journal of infectious diseases,* 208**,** 1335-41.

FORSLUND, O., JOHANSSON, H., MADSEN, K. G. & KOFOED, K. 2013c. The nasal mucosa contains a large spectrum of human papillomavirus types from the Betapapillomavirus and Gammapapillomavirus genera. *J Infect Dis,* 208**,** 1335-41.

FOULONGNE, V., SAUVAGE, V., HEBERT, C., DEREURE, O., CHEVAL, J., GOUILH, M. A., PARIENTE, K., SEGONDY, M., BURGUIÈRE, A., MANUGUERRA, J.-C., CARO, V. & ELOIT, M. 2012. Human Skin Microbiota: High Diversity of DNA Viruses Identified on the Human Skin by High Throughput Sequencing. *PLOS ONE,* 7**,** e38499.

GARCIA-VALLVE, S., ALONSO, A. & BRAVO, I. G. 2005. Papillomaviruses: different genes have different histories. *Trends Microbiol,* 13**,** 514-21.

GARCÍA-VALLVÉ, S., ALONSO, Á. & BRAVO, I. G. 2005. Papillomaviruses: different genes have different histories. *Trends in Microbiology,* 13**,** 514-521.

GHAI, J., OSTROW, R. S., TOLAR, J., MCGLENNEN, R. C., LEMKE, T. D., TOBOLT, D., LIU, Z. & FARAS, A. J. 1996. The E5 gene product of rhesus papillomavirus is an activator of endogenous Ras and phosphatidylinositol-3'-kinase in NIH 3T3 cells. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 12879-12884.

GIBBS, M. J., ARMSTRONG, J. S. & GIBBS, A. J. 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics,* 16**,** 573-582.

GIULIANO, A. R., TORTOLERO-LUNA, G., FERRER, E., BURCHELL, A. N., DE SANJOSE, S., KJAER, S. K., MUNOZ, N., SCHIFFMAN, M. & BOSCH, F. X. 2008. Epidemiology of human papillomavirus infection in men, cancers other than cervical and benign conditions. *Vaccine***,** 2008 Aug 19;26 Suppl 10:K17-28.

GOON, P., SONNEX, C., JANI, P., STANLEY, M. & SUDHOFF, H. 2008. Recurrent respiratory papillomatosis: an overview of current thinking and treatment. *European Archives of Oto-Rhino-Laryngology,* 265**,** 147-151.

GRACE, M. & MUNGER, K. 2017. Proteomic analysis of the gamma human papillomavirus type 197 E6 and E7 associated cellular proteins. *Virology,* 500**,** 71-81.

GRAHAM, S. V. 2016. Human Papillomavirus E2 Protein: Linking Replication, Transcription, and RNA Processing. *Journal of Virology,* 90**,** 8384-8388.

GRAHAM, SHEILA V. 2017. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clinical Science,* 131**,** 2201-2221.

GRCE, M. & MRAVAK-STIPETIĆ, M. 2014. Human papillomavirus–associated diseases. *Clinics in Dermatology,* 32**,** 253-258.

GUINDON, S., DUFAYARD, J. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology,* 59**,** 307-321.

HAGENSEE, M. E., OLSON, N. H., BAKER, T. S. & GALLOWAY, D. A. 1994. Three-dimensional structure of vaccinia virus-produced human papillomavirus type 1 capsids. *J Virol,* 68**,** 4503-5.

HANDISURYA, A., DAY, P. M., THOMPSON, C. D., BUCK, C. B., KWAK, K., RODEN, R. B., LOWY, D. R. & SCHILLER, J. T. 2012. Murine skin and vaginal mucosa are similarly susceptible to infection by pseudovirions of different papillomavirus classifications and species. *Virology,* 433**,** 385-94.

HARARI, A., CHEN, Z. & BURK, R. D. 2014. Human papillomavirus genomics: past, present and future. *Curr Probl Dermatol,* 45**,** 1-18.

HIGHAM, T., DOUKA, K., WOOD, R., RAMSEY, C. B., BROCK, F., BASELL, L., CAMPS, M., ARRIZABALAGA, A., BAENA, J., BARROSO-RUIZ, C., BERGMAN, C., BOITARD, C., BOSCATO, P., CAPARROS, M., CONARD, N. J., DRAILY, C., FROMENT, A., GALVAN, B., GAMBASSINI, P., GARCIA-MORENO, A., GRIMALDI, S., HAESAERTS, P., HOLT, B., IRIARTE-CHIAPUSSO, M. J., JELINEK, A., JORDA PARDO, J. F., MAILLO-FERNANDEZ, J. M., MAROM, A., MAROTO, J., MENENDEZ, M., METZ, L., MORIN, E., MORONI, A., NEGRINO, F., PANAGOPOULOU, E., PERESANI, M., PIRSON, S., DE LA RASILLA, M., RIEL-SALVATORE, J., RONCHITELLI, A., SANTAMARIA, D., SEMAL, P., SLIMAK, L., SOLER, J., SOLER, N., VILLALUENGA, A., PINHASI, R. & JACOBI, R. 2014. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature,* 512**,** 306-9.

HIROCHIKA, H., HIROCHIKA, R., BROKER, T. R. & CHOW, L. T. 1988. Functional mapping of the human papillomavirus type 11 transcriptional enhancer and its interaction with the trans-acting E2 proteins. *Genes Dev,* 2**,** 54-67.

HIROSE, Y., ONUKI, M., TENJIMBAYASHI, Y., MORI, S., ISHII, Y., TAKEUCHI, T., TASAKA, N., SATOH, T., MORISADA, T., IWATA, T., MIYAMOTO, S., MATSUMOTO, K., SEKIZAWA, A. & KUKIMOTO, I. 2018. Within-Host Variations of Human Papillomavirus Reveal APOBEC Signature Mutagenesis in the Viral Genome. *J Virol,* 92.

IANNACONE, M. R., GHEIT, T., WATERBOER, T., GIULIANO, A. R., MESSINA, J. L., FENSKE, N. A., CHERPELIS, B. S., SONDAK, V. K., ROETZHEIM, R. G., FERRER-GIL, S., MICHAEL, K. M., MCKAY-CHOPIN, S., PAWLITA, M., TOMMASINO, M. & ROLLISON, D. E. 2013. Case-control study of cutaneous human papillomavirus infection in Basal cell carcinoma of the skin. *J Invest Dermatol,* 2013 Jun;133(6):1512-20. doi**,** 10.1038/jid.2012.478.

IFTNER, T., HAEDICKE-JARBOUI, J., WU, S.-Y. & CHIANG, C.-M. 2017. Involvement of Brd4 in different steps of the papillomavirus life cycle. *Virus Research,* 231**,** 76-82.

JAMES, C. & ROBERTS, S. 2016. Viral Interactions with PDZ Domain-Containing Proteins—An Oncogenic Trait? *Pathogens,* 5**,** 8.

KALYAANAMOORTHY, S., MINH, B. Q., WONG, T. K. F., VON HAESELER, A. & JERMIIN, L. S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods,* 14**,** 587-589.

KEELE, B. F., GIORGI, E. E., SALAZAR-GONZALEZ, J. F., DECKER, J. M., PHAM, K. T., SALAZAR, M. G., SUN, C., GRAYSON, T., WANG, S., LI, H., WEI, X., JIANG, C., KIRCHHERR, J. L., GAO, F., ANDERSON, J. A., PING, L. H., SWANSTROM, R., TOMARAS, G. D., BLATTNER, W. A., GOEPFERT, P. A., KILBY, J. M., SAAG, M. S., DELWART, E. L., BUSCH, M. P., COHEN, M. S., MONTEFIORI, D. C., HAYNES, B. F., GASCHEN, B., ATHREYA, G. S., LEE, H. Y., WOOD, N., SEOIGHE, C., PERELSON, A. S., BHATTACHARYA, T., KORBER, B. T., HAHN, B. H. & SHAW, G. M. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A,* 105**,** 7552-7.

KENNEDY, M. B. 1995. Origin of PDZ (DHR, GLGF) domains. *Trends Biochem Sci,* 20**,** 350.

KING, E. M., GILSON, R., BEDDOWS, S., SOLDAN, K., PANWAR, K., YOUNG, C., PRAH, P., JIT, M., EDMUNDS, W. J. & SONNENBERG, P. 2015. Human papillomavirus DNA in men who have sex with men: type-specific prevalence, risk factors and implications for vaccination strategies. *Br J Cancer,* 112**,** 1585-93.

KISHINO, H. & HASEGAWA, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution,* 29**,** 170-179.

KISHINO, H., MIYATA, T. & HASEGAWA, M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution,* 31**,** 151-160.

KOHLER, A., GOTTSCHLING, M., MANNING, K., LEHMANN, M. D., SCHULZ, E., KRUGER-CORCORAN, D., STOCKFLETH, E. & NINDL, I. 2011. Genomic characterization of ten novel cutaneous human papillomaviruses from keratotic lesions of immunosuppressed patients. *J Gen Virol,* 92**,** 1585-94.

KORONA, D. A., LECOMPTE, K. G. & PURSELL, Z. F. 2011. The high fidelity and unique error signature of human DNA polymerase epsilon. *Nucleic Acids Res,* 39**,** 1763-73.

LA COUR, T., KIEMER, L., MOLGAARD, A., GUPTA, R., SKRIVER, K. & BRUNAK, S. 2004. Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel,* 17**,** 527-36.

LAITY, J. H., LEE, B. M. & WRIGHT, P. E. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol,* 11**,** 39-46.

LANGE, A., MILLS, R. E., LANGE, C. J., STEWART, M., DEVINE, S. E. & CORBETT, A. H. 2007. Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem,* 282**,** 5101-5.

LANIOSZ, V., DABYDEEN, S. A., HAVENS, M. A. & MENESES, P. I. 2009. Human papillomavirus type 16 infection of human keratinocytes requires clathrin and caveolin-1 and is brefeldin a sensitive. *J Virol,* 83**,** 8221-32.

LATSUZBAIA, A., ARBYN, M., DUTTA, S., FISCHER, M., GHEIT, T., TAPP, J., TOMMASINO, M., WEYERS, S. & MOSSONG, J. 2018. Complete Genome Sequence of a Novel Human Gammapapillomavirus Isolated from a Cervical Swab in Luxembourg. *Genome Announc,* 6.

LAZARCZYK, M., CASSONNET, P., PONS, C., JACOB, Y. & FAVRE, M. 2009. The EVER Proteins as a Natural Barrier against Papillomaviruses: a New Insight into the Pathogenesis of Human Papillomavirus Infections. *Microbiology and Molecular Biology Reviews,* 73**,** 348-370.

LEECHANACHAI, P., BANKS, L., MOREAU, F. & MATLASHEWSKI, G. 1992. The E5 gene from human papillomavirus type 16 is an oncogene which enhances growth factor-mediated signal transduction to the nucleus. *Oncogene,* 7**,** 19-25.

LETUNIC, I. & BORK, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res,* 44**,** W242-5.

LI, J., CAI, H., XU, Z., WANG, Q., HANG, D., SHEN, N., LIU, M., ZHANG, C., ABLIZ, A. & KE, Y. 2012. Nine complete genome sequences of cutaneous human papillomavirus genotypes

isolated from healthy skin of individuals living in rural He Nan province, China. *J Virol,* 86**,** 11936.

LI, L., BARRY, P., YEH, E., GLASER, C., SCHNURR, D. & DELWART, E. 2009. Identification of a novel human gammapapillomavirus species. *J Gen Virol,* 90**,** 2413-7.

LI, R., KNIGHT, J., BREAM, G., STENLUND, A. & BOTCHAN, M. 1989. Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev,* 3**,** 510-26.

LI, W., COWLEY, A., ULUDAG, M., GUR, T., MCWILLIAM, H., SQUIZZATO, S., PARK, Y. M., BUSO, N. & LOPEZ, R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res,* 43**,** W580-4.

LISSOUBA, P., VAN DE PERRE, P. & AUVERT, B. 2013. Association of genital human papillomavirus infection with HIV acquisition: a systematic review and meta-analysis. *Sex Transm Infect,* 89**,** 350-6.

LIU, W. J., GISSMANN, L., SUN, X. Y., KANJANAHALUETHAI, A., MULLER, M., DOORBAR, J. & ZHOU, J. 1997. Sequence close to the N-terminus of L2 protein is displayed on the surface of bovine papillomavirus type 1 virions. *Virology,* 227**,** 474-83.

LONGWORTH, M. S. & LAIMINS, L. A. 2004. Pathogenesis of Human Papillomaviruses in Differentiating Epithelia. *Microbiology and Molecular Biology Reviews,* 68**,** 362-372.

LU, B., WU, Y., NIELSON, C. M., FLORES, R., ABRAHAMSEN, M., PAPENFUSS, M., HARRIS, R. B. & GIULIANO, A. R. 2009. Factors associated with acquisition and clearance of human papillomavirus infection in a cohort of US men: a prospective study. *J Infect Dis,* 199**,** 362-71.

MA, Y., MADUPU, R., KARAOZ, U., NOSSA, C. W., YANG, L., YOOSEPH, S., YACHIMSKI, P. S., BRODIE, E. L., NELSON, K. E. & PEI, Z. 2014a. Human Papillomavirus Community in Healthy Persons, Defined by Metagenomics Analysis of Human Microbiome Project Shotgun Sequencing Data Sets. *Journal of Virology,* 88**,** 4786-4797.

MA, Y., MADUPU, R., KARAOZ, U., NOSSA, C. W., YANG, L., YOOSEPH, S., YACHIMSKI, P. S., BRODIE, E. L., NELSON, K. E. & PEI, Z. 2014b. Human papillomavirus community in healthy persons, defined by metagenomics analysis of human microbiome project shotgun sequencing data sets. *Journal of virology,* 88**,** 4786-97.

MAGLENNON, G. A., MCINTOSH, P. & DOORBAR, J. 2011. Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression. *Virology,* 414**,** 153-163.

MARRA, E., LIN, C. & CLIFFORD, G. M. 2018. Type-specific anal human papillomavirus prevalence among men, according to sexual preference and HIV status: a systematic literature review and meta-analysis. *J Infect Dis*.

MARTIN, D. & RYBICKI, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics,* 16**,** 562-563.

MARTIN, D. P., POSADA, D., CRANDALL, K. A. & WILLIAMSON, C. 2005a. A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. *AIDS Research and Human Retroviruses,* 21**,** 98-102.

MARTIN, D. P., WILLIAMSON, C. & POSADA, D. 2005b. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics,* 21**,** 260-262.

MARTIN, E., DANG, J., BZHALAVA, D., STERN, J., EDELSTEIN, Z. R., KOUTSKY, L. A., KIVIAT, N. B. & FENG, Q. 2014. Characterization of three novel human papillomavirus types isolated from oral rinse samples of healthy individuals. *J Clin Virol,* 59**,** 30-7.

MARTINEZ-ZAPIEN, D., RUIZ, F. X., POIRSON, J., MITSCHLER, A., RAMIREZ-RAMOS, J., FORSTER, A., COUSIDO-SIAH, A., MASSON, M., POL, S. V., PODJARNY, A., TRAVÉ, G. & ZANIER, K. 2016. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature,* 529**,** 541-545.

MBULAWA, Z. Z., COETZEE, D., MARAIS, D. J., KAMUPIRA, M., ZWANE, E., ALLAN, B., CONSTANT, D., MOODLEY, J. R., HOFFMAN, M. & WILLIAMSON, A. L. 2009. Genital human papillomavirus prevalence and human papillomavirus concordance in heterosexual couples are positively associated with human immunodeficiency virus coinfection. *J Infect Dis,* 199(10)**,** 1514-1524.

MBULAWA, Z. Z., MARAIS, D. J., JOHNSON, L. F., BOULLE, A., COETZEE, D. & WILLIAMSON, A. L. 2010. Influence of human immunodeficiency virus and CD4 count on the prevalence of human papillomavirus in heterosexual couples. *J Gen Virol,* 2010 Dec;91(Pt 12):3023-31. doi**,** 10.1099/vir.0.020669-0.

MBULAWA, Z. Z., MARAIS, D. J., JOHNSON, L. F., COETZEE, D. & WILLIAMSON, A. L. 2012. Impact of human immunodeficiency virus on the natural history of human papillomavirus genital infection in South African men and women. *J Infect Dis,* 2012 Jul 1;206(1):15-27. doi**,** 10.1093/infdis/jis299.

MCBRIDE, A. A. 2013. The papillomavirus E2 proteins. *Virology,* 445**,** 57-79.

MCBRIDE, A. A. 2017. Mechanisms and strategies of papillomavirus replication. *Biol Chem,* 398**,** 919-927.

MCLAUGHLIN-DRUBIN, M. E. & MUNGER, K. 2009a. The human papillomavirus E7 oncoprotein. *Virology,* 384**,** 335-44.

MCLAUGHLIN-DRUBIN, M. E. & MUNGER, K. 2009b. Oncogenic activities of human papillomaviruses. *Virus Res,* 143(2). **,** 195-208.

MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res,* 41**,** W597-600.

MEIRING, T. L., MBULAWA, Z. Z. A., LESOSKY, M., COETZEE, D. & WILLIAMSON, A. L. 2017. High diversity of alpha, beta and gamma human papillomaviruses in genital samples from HIV-negative and HIV-positive heterosexual South African men. *Papillomavirus Res,* 3**,** 160-167.

MEIRING, T. L., SALIMO, A. T., COETZEE, B., MAREE, H. J., MOODLEY, J., HITZEROTH, II, FREEBOROUGH, M. J., RYBICKI, E. P. & WILLIAMSON, A. L. 2012. Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits. *Virol J*, 2012 Aug 16;9:164.

MEISAL, R., ROUNGE, T. B., CHRISTIANSEN, I. K., EIELAND, A. K., WORREN, M. M., MOLDEN, T. F., KOMMEDAL, O., HOVIG, E., LEEGAARD, T. M. & AMBUR, O. H. 2017. HPV Genotyping of Modified General Primer-Amplicons Is More Analytically Sensitive and Specific by Sequencing than by Hybridization. *PLoS One,* 12**,** e0169074.

MILLER, D. L., PURICELLI, M. D. & STACK, M. S. 2012. Virology and molecular pathogenesis of HPV (human papillomavirus)-associated oropharyngeal squamous cell carcinoma. *Biochem J,* 443**,** 339-53.

MINH, B. Q., NGUYEN, M. A. T. & VON HAESELER, A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution,* 30**,** 1188-1195.

MIRABELLO, L., YEAGER, M., YU, K., CLIFFORD, G. M., XIAO, Y., ZHU, B., CULLEN, M., BOLAND, J. F., WENTZENSEN, N., NELSON, C. W., RAINE-BENNETT, T., CHEN, Z., BASS, S., SONG, L., YANG, Q., STEINBERG, M., BURDETT, L., DEAN, M., ROBERSON, D., MITCHELL, J., LOREY, T., FRANCESCHI, S., CASTLE, P. E., WALKER, J., ZUNA, R., KREIMER, A. R., BEACHLER, D. C., HILDESHEIM, A., GONZALEZ, P., PORRAS, C., BURK, R. D. & SCHIFFMAN, M. 2017. HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell,* 170**,** 1164-1174.e6.

MITSUISHI, T., OHSAWA, I., KATO, T., EGAWA, N. & KIYONO, T. 2013. Molecular cloning and characterisation of a novel type of human papillomavirus 160 isolated from a flat wart of an immunocompetent patient. *PLoS One,* 8**,** e79592.

MLAKAR, B., KOCJAN, B. J., HOSNJAK, L., FUJS KOMLOS, K., MILOSEVIC, M. & POLJAK, M. 2014. Betapapillomaviruses in the anal canal of HIV positive and HIV negative men who have sex with men. *J Clin Virol,* 61**,** 237-41.

MODIS, Y., TRUS, B. L. & HARRISON, S. C. 2002. Atomic model of the papillomavirus capsid. *Embo J,* 21**,** 4754-62.

MOODY, C. A. & LAIMINS, L. A. 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer,* 10**,** 550-60.

MOSCICKI, A.-B., MA, Y., GHEIT, T., MCKAY-CHOPIN, S., FARHAT, S., WIDDICE, L. E. & TOMMASINO, M. 2017. Prevalence and Transmission of Beta and Gamma Human Papillomavirus in Heterosexual Couples. *Open Forum Infectious Diseases,* 4**,** ofw216-ofw216.

MÜHR, L. S. A., EKLUND, C. & DILLNER, J. 2018. Towards quality and order in human papillomavirus research. *Virology,* 519**,** 74-76.

MULLER, E. E., REBE, K., CHIRWA, T. F., STRUTHERS, H., MCINTYRE, J. & LEWIS, D. A. 2016. The prevalence of human papillomavirus infections and associated risk factors in men-who-have-sex-with-men in Cape Town, South Africa. *BMC Infect Dis,* 16**,** 440.

MUNGER, K. 2002. The role of human papillomaviruses in human cancers. *Front Biosci,* 1**,** d641-9.

MUNGER, K., BALDWIN, A., EDWARDS, K. M., HAYAKAWA, H., NGUYEN, C. L., OWENS, M., GRACE, M. & HUH, K. 2004. Mechanisms of human papillomavirus-induced oncogenesis. *J Virol,* 78**,** 11451-60.

MURAHWA, A. T., MEIRING, T. L., MBULAWA, Z. Z. A. & WILLIAMSON, A. L. 2018. Complete Genome Sequences of Four Novel Human Gammapapillomavirus Types, HPV-219, HPV-220, HPV-221, and HPV-222, Isolated from Penile Skin Swabs from South African Men. *Genome Announc,* 6.

MURAHWA, A. T., MUCHEMWA, F. C., DURI, K., KANYERA, R. B., TSHABALALA, M., MANHANZVA, M. T., MAPINGURE, M. P. & STRAY-PEDERSEN, B. 2014. Frequency of Betapapillomavirus infections among HIV infected and uninfected Black Zimbabweans with cutaneous lesions. *J Med Virol,* 87**,** 478-84.

NARECHANIA, A., CHEN, Z., DESALLE, R. & BURK, R. D. 2005a. Phylogenetic Incongruence among Oncogenic Genital Alpha Human Papillomaviruses. *Journal of Virology,* 79**,** 15503-15510.

NARECHANIA, A., CHEN, Z., DESALLE, R. & BURK, R. D. 2005b. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol,* 79**,** 15503-10.

NEI, M. & GOJOBORI, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol,* 3**,** 418-26.

NELSON, L. M., ROSE, R. C., LEROUX, L., LANE, C., BRUYA, K. & MOROIANU, J. 2000. Nuclear import and DNA binding of human papillomavirus type 45 L1 capsid protein. *J Cell Biochem,* 79**,** 225-38.

NEWHOUSE, C. D. & SILVERSTEIN, S. J. 2001. Orientation of a novel DNA binding site affects human papillomavirus-mediated transcription and replication. *J Virol,* 75**,** 1722-1735.

NIELSON, C. M., HARRIS, R. B., DUNNE, E. F., ABRAHAMSEN, M., PAPENFUSS, M. R., FLORES, R., MARKOWITZ, L. E. & GIULIANO, A. R. 2007. Risk factors for anogenital human papillomavirus infection in men. *J Infect Dis,* 196**,** 1137-45.

NIELSON, C. M., HARRIS, R. B., FLORES, R., ABRAHAMSEN, M., PAPENFUSS, M. R., DUNNE, E. F., MARKOWITZ, L. E. & GIULIANO, A. R. 2009. Multiple-type human papillomavirus infection in male anogenital sites: prevalence and associated factors. *Cancer Epidemiol Biomarkers Prev,* 18**,** 1077-83.

NILYANIMIT, P., CHANSAENROJ, J., POOMIPAK, W., PRAIANANTATHAVORN, K., PAYUNGPORN, S. & POOVORAWAN, Y. 2018. Comparison of Four Human Papillomavirus Genotyping Methods: Next-generation Sequencing, INNO-LiPA, Electrochemical DNA Chip, and Nested-PCR. *Ann Lab Med,* 38**,** 139-146.

NOBRE, R. J., HERRAEZ-HERNANDEZ, E., FEI, J. W., LANGBEIN, L., KADEN, S., GRONE, H. J. & DE VILLIERS, E. M. 2009a. E7 oncoprotein of novel human papillomavirus type 108 lacking the E6 gene induces dysplasia in organotypic keratinocyte cultures. *Journal of virology,* 83**,** 2907-16.

NOBRE, R. J., HERRAEZ-HERNANDEZ, E., FEI, J. W., LANGBEIN, L., KADEN, S., GRONE, H. J. & DE VILLIERS, E. M. 2009b. E7 oncoprotein of novel human papillomavirus type 108 lacking the E6 gene induces dysplasia in organotypic keratinocyte cultures. *J Virol,* 83**,** 2907-16.

OLESEN, T. B., MWAISELAGE, J., IFTNER, T., KAHESA, C., RASCH, V., FREDERIKSEN, K., MUNK, C. & KJAER, S. K. 2017. Risk factors for genital human papillomavirus among men in Tanzania. *J Med Virol,* 89**,** 345-351.

OSTRBENK, A., KOCJAN, B. J., HOSNJAK, L., LI, J., DENG, Q., STERBENC, A. & POLJAK, M. 2015. Identification of a Novel Human Papillomavirus, Type HPV199, Isolated from a Nasopharynx and Anal Canal, and Complete Genomic Characterization of Papillomavirus Species Gamma-12. *PLoS One,* 10**,** e0138628.

PADIDAM, M., SAWYER, S. & FAUQUET, C. M. 1999. Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology,* 265**,** 218-225.

PEH, W. L., MIDDLETON, K., CHRISTENSEN, N., NICHOLLS, P., EGAWA, K., SOTLAR, K., BRANDSMA, J., PERCIVAL, A., LEWIS, J., LIU, W. J. & DOORBAR, J. 2002. Life Cycle Heterogeneity in Animal Models of Human Papillomavirus-Associated Disease. *Journal of Virology,* 76**,** 10401-10416.

PELKMANS, L. & HELENIUS, A. 2003. Insider information: what viruses tell us about endocytosis. *Curr Opin Cell Biol,* 15**,** 414-22.

PIM, D., COLLINS, M. & BANKS, L. 1992. Human papillomavirus type 16 E5 gene stimulates the transforming activity of the epidermal growth factor receptor. *Oncogene,* 7**,** 27-32.

PINIDIS, P., TSIKOURAS, P., IATRAKIS, G., ZERVOUDIS, S., KOUKOULI, Z., BOTHOU, A., GALAZIOS, G. & VLADAREANU, S. 2016. Human Papilloma Virus' Life Cycle and Carcinogenesis. *Maedica (Buchar),* 11**,** 48-54.

POSADA, D. & CRANDALL, K. A. 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences,* 98**,** 13757-13762.

RAMBAUT, A., DRUMMOND, A. J., XIE, D., BAELE, G. & SUCHARD, M. A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology,* 67**,** 901-904.

RECTOR, A., LEMEY, P., TACHEZY, R., MOSTMANS, S., GHIM, S. J., VAN DOORSLAER, K., ROELKE, M., BUSH, M., MONTALI, R. J., JOSLIN, J., BURK, R. D., JENSON, A. B., SUNDBERG, J. P., SHAPIRO, B. & VAN RANST, M. 2007. Ancient papillomavirus-host co-speciation in Felidae. *Genome Biol,* 8**,** R57.

REESE, M. G., EECKMAN, F. H., KULP, D. & HAUSSLER, D. 1997. Improved splice site detection in Genie. *J Comput Biol,* 4**,** 311-23.

REPP, K. K., NIELSON, C. M., FU, R., SCHAFER, S., LAZCANO-PONCE, E., SALMERON, J., QUITERIO, M., VILLA, L. L. & GIULIANO, A. R. 2012. Male human papillomavirus prevalence and association with condom use in Brazil, Mexico, and the United States. *J Infect Dis,* 205**,** 1287-93.

RHO, J., DE VILLIERS, E. M. & CHOE, J. 1996. Transforming activities of human papillomavirus type 59 E5, E6 and E7 open reading frames in mouse C127 cells. *Virus Res,* 44**,** 57-65.

ROBERTS, S., HILLMAN, M. L., KNIGHT, G. L. & GALLIMORE, P. H. 2003. The ND10 Component Promyelocytic Leukemia Protein Relocates to Human Papillomavirus Type 1 E4

Intranuclear Inclusion Bodies in Cultured Keratinocytes and in Warts. *Journal of Virology,* 77**,** 673-684.

ROCHA RODRÍGUEZ, M. D. R., JUÁREZ JUÁREZ, M. E., RUIZ JIMÉNEZ, M. M., RAMÍREZ BANDA, X. G., GAYTÁN SANCHÉZ, M. D. R. & CONTRERAS VALERO, P. 2012. Identificación de factores de riesgo para contraer virus del papiloma humano en sexoservidoras. *Revista Cubana de Obstetricia y Ginecología,* 38**,** 244-255.

RODEN, R. B. S. & STERN, P. L. 2018. Opportunities and challenges for human papillomavirus vaccination in cancer. *Nature Reviews Cancer,* 18**,** 240.

RODRIGO, A. G., LEARN JR., GERALD H. 2000. HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences. *Dordrecht, Netherlands: Kluwer Academic Publishers.,* Chapter 4**,** pages 55-72.

RODRIGUEZ-ALVAREZ, M. I., GOMEZ-URQUIZA, J. L., HUSEIN-EL AHMED, H., ALBENDIN-GARCIA, L., GOMEZ-SALGADO, J. & CANADAS-DE LA FUENTE, G. A. 2018. Prevalence and Risk Factors of Human Papillomavirus in Male Patients: A Systematic Review and Meta-Analysis. *Int J Environ Res Public Health,* 15.

ROMAN, A. & MUNGER, K. 2013a. The papillomavirus E7 proteins. *Virology,* 445**,** 138-168.

ROMAN, A. & MUNGER, K. 2013b. The papillomavirus E7 proteins. *Virology,* 445**,** 138-68.

RONCO, G., DILLNER, J., ELFSTRÖM, K. M., TUNESI, S., SNIJDERS, P. J. F., ARBYN, M., KITCHENER, H., SEGNAN, N., GILHAM, C., GIORGI-ROSSI, P., BERKHOF, J., PETO, J. & MEIJER, C. J. L. M. 2014. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *The Lancet,* 383**,** 524-532.

ROSITCH, A. F., HUDGENS, M. G., BACKES, D. M., MOSES, S., AGOT, K., NYAGAYA, E., SNIJDERS, P. J., MEIJER, C. J., BAILEY, R. C. & SMITH, J. S. 2012. Vaccine-relevant human papillomavirus (HPV) infections and future acquisition of high-risk HPV types in men. *J Infect Dis,* 206**,** 669-77.

RUBIO, I., SEITZ, H., CANALI, E., SEHR, P., BOLCHI, A., TOMMASINO, M., OTTONELLO, S. & MULLER, M. 2011. The N-terminal region of the human papillomavirus L2 protein contains overlapping binding sites for neutralizing, cross-neutralizing and non-neutralizing antibodies. *Virology,* 409**,** 348-59.

SCHIFFMAN, M., HERRERO, R., DESALLE, R., HILDESHEIM, A., WACHOLDER, S., CECILIA RODRIGUEZ, A., BRATTI, M. C., SHERMAN, M. E., MORALES, J., GUILLEN, D., ALFARO, M., HUTCHINSON, M., WRIGHT, T. C., SOLOMON, D., CHEN, Z., SCHUSSLER, J., CASTLE, P. E. & BURK, R. D. 2005a. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology,* 337**,** 76-84.

SCHIFFMAN, M., HERRERO, R., DESALLE, R., HILDESHEIM, A., WACHOLDER, S., RODRIGUEZ, A. C., BRATTI, M. C., SHERMAN, M. E., MORALES, J., GUILLEN, D., ALFARO, M., HUTCHINSON, M., WRIGHT, T. C., SOLOMON, D., CHEN, Z., SCHUSSLER, J., CASTLE, P. E. & BURK, R. D. 2005b. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology,* 337**,** 76-84.

SCHILLER, J. T., DAY, P. M. & KINES, R. C. 2010. Current understanding of the mechanism of HPV infection. *Gynecol Oncol,* 118**,** S12-7.

SCHILLER, J. T. & LOWY, D. R. 2012. Understanding and learning from the success of prophylactic human papillomavirus vaccines. *Nat Rev Microbiol,* 10**,** 681-92.

SELINKA, H. C., FLORIN, L., PATEL, H. D., FREITAG, K., SCHMIDTKE, M., MAKAROV, V. A. & SAPP, M. 2007. Inhibition of transfer to secondary receptors by heparan sulfate-binding drug or antibody induces noninfectious uptake of human papillomavirus. *J Virol,* 81**,** 10970-80.

SHAH, S. D., DOORBAR, J. & GOLDSTEIN, R. A. 2010. Analysis of Host–Parasite Incongruence in Papillomavirus Evolution Using Importance Sampling. *Molecular Biology and Evolution,* 27**,** 1301-1314.

SHIMODAIRA, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol,* 51**,** 492-508.

SHIMODAIRA, H. & HASEGAWA, M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution,* 16**,** 1114-1114.

SICHERO, L., NYITRAY, A. G., NUNES, E. M., NEPAL, B., FERREIRA, S., SOBRINHO, J. S., BAGGIO, M. L., GALAN, L., SILVA, R. C., LAZCANO-PONCE, E., GIULIANO, A. R. & VILLA, L. L. 2015a. Diversity of human papillomavirus in the anal canal of men: the HIM Study. *Clin Microbiol Infect,* 21**,** 502-9.

SICHERO, L., NYITRAY, A. G., NUNES, E. M., NEPAL, B., FERREIRA, S., SOBRINHO, J. S., BAGGIO, M. L., GALAN, L., SILVA, R. C., LAZCANO-PONCE, E., GIULIANO, A. R. & VILLA, L. L. 2015b. Diversity of human papillomavirus in the anal canal of men: the HIM Study. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases,* 21**,** 502-9.

SICHERO, L., PIERCE CAMPBELL, C. M., FERREIRA, S., SOBRINHO, J. S., LUIZA BAGGIO, M., GALAN, L., SILVA, R. C., LAZCANO-PONCE, E., GIULIANO, A. R. & VILLA, L. L. 2013. Broad HPV distribution in the genital region of men from the HPV infection in men (HIM) study. *Virology,* 443**,** 214-7.

SICHERO, L., PIERCE CAMPBELL, C. M., FULP, W., FERREIRA, S., SOBRINHO, J. S., BAGGIO, M., GALAN, L., SILVA, R. C., LAZCANO-PONCE, E., GIULIANO, A. R. & VILLA, L. L. 2014a. High genital prevalence of cutaneous human papillomavirus DNA on male genital skin: the HPV Infection in Men Study. *BMC Infect Dis,* 14**,** 677.

SICHERO, L., PIERCE CAMPBELL, C. M., FULP, W., FERREIRA, S., SOBRINHO, J. S., BAGGIO, M., GALAN, L., SILVA, R. C., LAZCANO-PONCE, E., GIULIANO, A. R. & VILLA, L. L. 2014b. High genital prevalence of cutaneous human papillomavirus DNA on male genital skin: the HPV Infection in Men Study. *BMC infectious diseases,* 14**,** 677.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M., SODING, J., THOMPSON, J. D. & HIGGINS, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol,* 7**,** 539.

SILVERBERG, M. J., LAU, B., JUSTICE, A. C., ENGELS, E., GILL, M. J., GOEDERT, J. J., KIRK, G. D., D'SOUZA, G., BOSCH, R. J., BROOKS, J. T., NAPRAVNIK, S., HESSOL, N. A., JACOBSON, L. P., KITAHATA, M. M., KLEIN, M. B., MOORE, R. D., RODRIGUEZ, B., ROURKE, S. B., SAAG,

M. S., STERLING, T. R., GEBO, K. A., PRESS, N., MARTIN, J. N. & DUBROW, R. 2012. Risk of anal cancer in HIV-infected and HIV-uninfected individuals in North America. *Clin Infect Dis,* 54**,** 1026-34.

SIMMONDS, P. 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J Gen Virol,* 96**,** 1193-206.

SIMMONDS, P., ADAMS, M. J., BENKO, M., BREITBART, M., BRISTER, J. R., CARSTENS, E. B., DAVISON, A. J., DELWART, E., GORBALENYA, A. E., HARRACH, B., HULL, R., KING, A. M., KOONIN, E. V., KRUPOVIC, M., KUHN, J. H., LEFKOWITZ, E. J., NIBERT, M. L., ORTON, R., ROOSSINCK, M. J., SABANADZOVIC, S., SULLIVAN, M. B., SUTTLE, C. A., TESH, R. B., VAN DER VLUGT, R. A., VARSANI, A. & ZERBINI, F. M. 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nat Rev Microbiol,* 15**,** 161-168.

SMELOV, V., MUWONGE, R., SOKOLOVA, O., MCKAY-CHOPIN, S., EKLUND, C., KOMYAKOV, B. & GHEIT, T. 2018. Beta and gamma human papillomaviruses in anal and genital sites among men: prevalence and determinants. *Scientific Reports,* 8**,** 8241.

SMITH, J. L., CAMPOS, S. K. & OZBUN, M. A. 2007. Human papillomavirus type 31 uses a caveolin 1- and dynamin 2-mediated entry pathway for infection of human keratinocytes. *J Virol,* 81**,** 9922-31.

SMITH, J. M. 1992. Analyzing the mosaic structure of genes. *J Mol Evol,* 34**,** 126-9.

SMOLA, S. 2014. Human papillomaviruses and skin cancer. *Adv Exp Med Biol,* 810**,** 192-207.

SONGYANG, Z., FANNING, A. S., FU, C., XU, J., MARFATIA, S. M., CHISHTI, A. H., CROMPTON, A., CHAN, A. C., ANDERSON, J. M. & CANTLEY, L. C. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science,* 275**,** 73-7.

SORIA-CARRASCO, V., TALAVERA, G., IGEA, J. & CASTRESANA, J. 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. *Bioinformatics,* 23**,** 2954-6.

SPODEN, G., FREITAG, K., HUSMANN, M., BOLLER, K., SAPP, M., LAMBERT, C. & FLORIN, L. 2008. Clathrin- and caveolin-independent entry of human papillomavirus type 16--involvement of tetraspanin-enriched microdomains (TEMs). *PLoS One,* 3**,** e3313.

STANLEY, M. A. 2012. Epithelial Cell Responses to Infection with Human Papillomavirus. *Clinical Microbiology Reviews,* 25**,** 215-222.

STANLEY, M. A., PETT, M. R. & COLEMAN, N. 2007. HPV: from infection to cancer. *Biochem Soc Trans,* 35**,** 1456-60.

STRIMMER, K. & RAMBAUT, A. 2002. Inferring confidence sets of possibly misspecified gene trees. *Proc Biol Sci,* 269**,** 137-42.

STUREGARD, E., JOHANSSON, H., EKSTROM, J., HANSSON, B. G., JOHNSSON, A., GUSTAFSSON, E., DILLNER, J. & FORSLUND, O. 2013. Human papillomavirus typing in reporting of condyloma. *Sex Transm Dis,* 40**,** 123-9.

SUAREZ, I. & TRAVE, G. 2018. Structural Insights in Multifunctional Papillomavirus Oncoproteins. *Viruses,* 10**,** 37.

SVERDRUP, F. & KHAN, S. A. 1995. Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA. *J Virol,* 69**,** 1319-23.

SYRJANEN, S. 2018. Oral manifestations of human papillomavirus infections. *Eur J Oral Sci,* 126 Suppl 1**,** 49-66.

TABERNA, M., MENA, M., PAVON, M. A., ALEMANY, L., GILLISON, M. L. & MESIA, R. 2017. Human papillomavirus-related oropharyngeal cancer. *Ann Oncol,* 28**,** 2386-2398.

TANTIN, D. 2013. Oct transcription factors in development and stem cells: insights and mechanisms. *Development,* 140**,** 2857-66.

TERAI, M. & BURK, R. D. 2002. Identification and characterization of 3 novel genital human papillomaviruses by overlapping polymerase chain reaction: candHPV89, candHPV90, and candHPV91. *J Infect Dis,* 185**,** 1794-7.

TITOLO, S., PELLETIER, A., SAUVE, F., BRAULT, K., WARDROP, E., WHITE, P. W., AMIN, A., CORDINGLEY, M. G. & ARCHAMBAULT, J. 1999. Role of the ATP-binding domain of the human papillomavirus type 11 E1 helicase in E2-dependent binding to the origin. *J Virol,* 73**,** 5282-93.

TOBIAN, A. A., GRABOWSKI, M. K., KIGOZI, G., REDD, A. D., EATON, K. P., SERWADDA, D., CORNISH, T. C., NALUGODA, F., WATYA, S., BUWEMBO, D., NKALE, J., WAWER, M. J., QUINN, T. C. & GRAY, R. H. 2013. Human papillomavirus clearance among males is associated with HIV acquisition and increased dendritic cell density in the foreskin. *J Infect Dis,* 207**,** 1713-22.

TOMAIĆ, V. 2016. Functional Roles of E6 and E7 Oncoproteins in HPV-Induced Malignancies at Diverse Anatomical Sites. *Cancers,* 8**,** 95.

TOMMASINO, M. 2017. The biology of beta human papillomaviruses. *Virus Research,* 231**,** 128-138.

TORRES, M., GHEIT, T., MCKAY-CHOPIN, S., RODRIGUEZ, C., ROMERO, J. D., FILOTICO, R., DONA, M. G., ORTIZ, M. & TOMMASINO, M. 2015. Prevalence of beta and gamma human papillomaviruses in the anal canal of men who have sex with men is influenced by HIV status. *J Clin Virol,* 67**,** 47-51.

TRIFINOPOULOS, J., NGUYEN, L. T., VON HAESELER, A. & MINH, B. Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res,* 44**,** W232-5.

URE, A. E. & FORSLUND, O. 2014a. Characterization of human papillomavirus type 154 and tissue tropism of gammapapillomaviruses. *PLoS One,* 2014 Feb 13;9(2):e89342. doi**,** 10.1371/journal.pone.0089342.

URE, A. E. & FORSLUND, O. 2014b. Characterization of human papillomavirus type 154 and tissue tropism of gammapapillomaviruses. *PLoS One,* 9**,** e89342.

VAN DE POEL, S., DREER, M., VELIC, A., MACEK, B., BASKARAN, P., IFTNER, T. & STUBENRAUCH, F. 2018. Identification and Functional Characterization of Phosphorylation Sites of the Human Papillomavirus 31 E8^E2 Protein. *J Virol,* 92.

VAN DOORSLAER, K. 2013. Evolution of the papillomaviridae. *Virology,* 445**,** 11-20.

VAN DOORSLAER, K., CHEN, Z., BERNARD, H. U., CHAN, P. K. S., DESALLE, R., DILLNER, J., FORSLUND, O., HAGA, T., MCBRIDE, A. A., VILLA, L. L., BURK, R. D. & ICTV REPORT, C. 2018. ICTV Virus Taxonomy Profile: Papillomaviridae. *J Gen Virol,* 99**,** 989-990.

VAN DOORSLAER, K., LI, Z., XIRASAGAR, S., MAES, P., KAMINSKY, D., LIOU, D., SUN, Q., KAUR, R., HUYEN, Y. & MCBRIDE, A. A. 2017a. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. *Nucleic Acids Res,* 45**,** 499-506.

VAN DOORSLAER, K. & MCBRIDE, A. A. 2016. Molecular archeological evidence in support of the repeated loss of a papillomavirus gene. *Sci Rep,* 6**,** 33028.

VAN DOORSLAER, K., RUOPPOLO, V., SCHMIDT, A., LESCROEL, A., JONGSOMJIT, D., ELROD, M., KRABERGER, S., STAINTON, D., DUGGER, K. M., BALLARD, G., AINLEY, D. G. & VARSANI, A. 2017b. Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. *Virus Evol,* 3**,** vex027.

VAN DOORSLAER, K., TAN, Q., XIRASAGAR, S., BANDARU, S., GOPALAN, V., MOHAMOUD, Y., HUYEN, Y. & MCBRIDE, A. A. 2013a. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic acids research,* 41**,** D571-578.

VAN DOORSLAER, K., TAN, Q., XIRASAGAR, S., BANDARU, S., GOPALAN, V., MOHAMOUD, Y., HUYEN, Y. & MCBRIDE, A. A. 2013b. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic acids research,* 41**,** D571-D578.

VANREGENMORTEL 2002. ICTV. *ictv report*.

VARDAS, E., GIULIANO, A. R., GOLDSTONE, S., PALEFSKY, J. M., MOREIRA, E. D., JR., PENNY, M. E., ARANDA, C., JESSEN, H., MOI, H., FERRIS, D. G., LIAW, K. L., MARSHALL, J. B., VUOCOLO, S., BARR, E., HAUPT, R. M., GARNER, E. I. & GURIS, D. 2011. External genital human papillomavirus prevalence and associated factors among heterosexual men on 5 continents. *J Infect Dis,* 203**,** 58-65.

VARSANI, A., VAN DER WALT, E., HEATH, L., RYBICKI, E. P., WILLIAMSON, A. L. & MARTIN, D. P. 2006. Evidence of ancient papillomavirus recombination. *J Gen Virol,* 87**,** 2527-31.

VARTANIAN, J. P., GUETARD, D., HENRY, M. & WAIN-HOBSON, S. 2008. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science,* 320**,** 230-3.

VENUTI, A., PAOLINI, F., NASIR, L., CORTEGGIO, A., ROPERTO, S., CAMPO, M. S. & BORZACCHIELLO, G. 2011. Papillomavirus E5: the smallest oncoprotein with many functions. *Mol Cancer,* 10**,** 140.

VÕSA, L., SUDAKOV, A., REMM, M., USTAV, M. & KURG, R. 2012. Identification and Analysis of Papillomavirus E2 Protein Binding Sites in the Human Genome. *Journal of Virology,* 86**,** 348-357.

WALLACE, N. A. & GALLOWAY, D. A. 2015. Novel Functions of the Human Papillomavirus E6 Oncoproteins. *Annu Rev Virol,* 2**,** 403-23.

WANG, J., ZHOU, D., PRABHU, A., SCHLEGEL, R. & YUAN, H. 2010. The canine papillomavirus and gamma HPV E7 proteins use an alternative domain to bind and destabilize the retinoblastoma protein. *PLoS Pathog,* 6**,** e1001089.

WANG, J. W. & RODEN, R. B. 2013. L2, the minor capsid protein of papillomavirus. *Virology,* 445**,** 175-86.

WANG, Q., GRIFFIN, H., SOUTHERN, S., JACKSON, D., MARTIN, A., MCINTOSH, P., DAVY, C., MASTERSON, P. J., WALKER, P. A., LASKEY, P., OMARY, M. B. & DOORBAR, J. 2004. Functional Analysis of the Human Papillomavirus Type 16 E1^ΛE4 Protein Provides a Mechanism for In Vivo and In Vitro Keratin Filament Reorganization. *Journal of Virology,* 78**,** 821-833.

WAYENGERA, M. 2012. Zinc finger arrays binding human papillomavirus types 16 and 18 genomic DNA: precursors of gene-therapeutics for in-situ reversal of associated cervical neoplasia. *Theor Biol Med Model,* 9**,** 30.

WEN, W., MEINKOTH, J. L., TSIEN, R. Y. & TAYLOR, S. S. 1995. Identification of a signal for rapid export of proteins from the nucleus. *Cell,* 82**,** 463-73.

WILLEMSEN, A. & BRAVO, I. G. 2018. Origin and evolution of papillomavirus (onco)genes and genomes. *bioRxiv*.

WILLIAMSON, A. L. 2015. The Interaction between Human Immunodeficiency Virus and Human Papillomaviruses in Heterosexuals in Africa. *J Clin Med,* 4**,** 579-92.

WISE-DRAPER, T. M. & WELLS, S. I. 2008. Papillomavirus E6 and E7 proteins and their cellular targets. *Front Biosci,* 13**,** 1003-17.

WOLF, M., GARCEA, R. L., GRIGORIEFF, N. & HARRISON, S. C. 2010. Subunit interactions in bovine papillomavirus. *Proc Natl Acad Sci U S A,* 107**,** 6298-303.

WOODBY, B., SCOTT, M. & BODILY, J. 2016. The Interaction Between Human Papillomaviruses and the Stromal Microenvironment. *Progress in molecular biology and translational science,* 144**,** 169-238.

WOODHAM, A. W., DA SILVA, D. M., SKEATE, J. G., RAFF, A. B., AMBROSO, M. R., BRAND, H. E., ISAS, J. M., LANGEN, R. & KAST, W. M. 2012. The S100A10 subunit of the annexin A2 heterotetramer facilitates L2-mediated human papillomavirus infection. *PLoS One,* 7**,** e43519.

ZAKRZEWSKA, K., REGALBUTO, E., PIERUCCI, F., ARVIA, R., MAZZOLI, S., GORI, A. & DE GIORGI, V. 2012. Pattern of HPV infection in basal cell carcinoma and in perilesional skin biopsies from immunocompetent patients. *Virol J***,** 2012 Dec 17;9:309.

ZANIER, K., CHARBONNIER, S., SIDI, A. O., MCEWEN, A. G., FERRARIO, M. G., POUSSIN-COURMONTAGNE, P., CURA, V., BRIMER, N., BABAH, K. O., ANSARI, T., MULLER, I., STOTE, R. H., CAVARELLI, J., VANDE POL, S. & TRAVE, G. 2013. Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. *Science,* 339**,** 694-8.

ZHOU, J., DOORBAR, J., SUN, X. Y., CRAWFORD, L. V., MCLEAN, C. S. & FRAZER, I. H. 1991. Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein. *Virology,* 185**,** 625-32.

ZHOU, J., STENZEL, D. J., SUN, X. Y. & FRAZER, I. H. 1993. Synthesis and assembly of infectious bovine papillomavirus particles in vitro. *J Gen Virol,* 74**,** 763-8.

ZOU, N., LIN, B. Y., DUAN, F., LEE, K. Y., JIN, G., GUAN, R., YAO, G., LEFKOWITZ, E. J., BROKER, T. R. & CHOW, L. T. 2000. The hinge of the human papillomavirus type 11 E2 protein contains major determinants for nuclear localization and nuclear matrix association. *J Virol,* 74**,** 3761-70.

ZUR HAUSEN, H. 2009. Papillomaviruses in the causation of human cancers - a brief historical account. *Virology,* 384**,** 260-5.